Email Security
Secure Email
**General Strategy**

- Basic scheme is pretty straight-forward
- Encrypt the message body with a symmetric cipher, using a randomly-generated traffic key
- Use public key cryptography to encrypt the traffic key to all recipients
- Digitally sign a hash of the message
- But there are many details
Some Details

- Obvious ones: which symmetric, public key, and hash algorithms to use?
- More subtle: which algorithms do the recipients understand?
- Where do certificates come from?
- Do you sign the plaintext or the ciphertext?
- How do you handle BCC?
- Will the ciphertext survive transit intact?
- How are header lines protected?
- What about attachments?
- Many possible answers to all of these questions
Transit Issues

- Not all mail systems accept all characters
- Very few are 8-bit clean
- Cryptographic transforms won’t survive even minor changes
- EBCDIC vs. ASCII? Unicode? Tabs versus blanks?
- Solution: encode all email in base 64, using characters all systems accept: A-Za-z0-9+/.
- Use 4 bytes to represent 3; overhead is 33%
- For padding, use = sign (see RFC 3548)
- Only those characters matter; everything else is deleted on receipt, including white space
Signing

- If you sign the plaintext and then encrypt, the sender’s identity is hidden from all except the proper recipients.
- If you sign the ciphertext, a gateway can verify signatures and present mail accordingly — perhaps better for anti-spam and anti-phishing.
Headers

- Headers change in transit
- Obvious example: `Received:` lines are added
- Less-obvious example: Email addresses are often rewritten to hide internal machines, and present clearer addresses to the outside:
  
  `smb@att.com → Steven.Bellovin@att.com`

- Consequence: headers are *not* protected by secure email schemes
- But — users look at (and search on) the headers
General Flow

- Collect input message
- Put in canonical form
- Encrypt and sign, or sign and encrypt
- Add metadata: encrypted traffic key, your certificate, algorithm identifiers, etc.
- Convert to transit form
- Embed in email message
Securing Transit

- Many pieces — but we can usually use TLS
- POP, IMAP, connection to submission server: all are by prearrangement
- Protect content; more important, protect passwords
- Problem area: road warriors vs. firewalls and anti-spam
1. Normal process: user composes mail on MUA; submits it to local submission server.
2. Optional internal hops
3. Outbound MTA contacts recipient’s MTA — interorganizational hop
4. Optional internal hops to recipient’s mail server (IMAP or POP)
5. IMAP or POP retrieval
6. How do we protect Step 3?
MTA to MTA Security

- Do we need to protect it at all?
- These are hard-to-tap links: phone company fiber, ISP backbones, etc.
- What about government wiretaps?
- Can use TLS — but what is the other side’s key? No PKI for Internet email!
- One answer: don’t worry; it’s still better than cleartext against passive eavesdroppers
- But — what about routing attacks?
Another reason to secure transit: *traffic analysis*

- Protect against traffic analysis — who is talking to whom
- Also: length, timing
- In practice, *extremely* valuable for law enforcement and intelligence agencies
- Less protected by US law
Secure Email

PGP and S/MIME

Approaches to Protecting Content
Certificate Style
Web of Trust
Does the Web of Trust Work?
Finding Public Keys
Which Style is Better?

Spam

Phishing

PGP and S/MIME
Approaches to Protecting Content

- Two major standards, PGP and S/MIME
- Many minor syntactic differences
- Major split by audience: computer scientists like PGP; mainstream users use S/MIME
- Biggest technical difference: how certificates are signed
Certificate Style

- S/MIME uses standard X.509 certificate format
- More importantly, X.509 certificates form a traditional PKI, with a root and a hierarchical structure
- Works well within an organization
- Between organizations, can work if it’s easy to find that organization’s root
- CU has no PKI — what is the PKI under which you’d find my cert? Why should you trust its root?
Web of Trust

- PGP use a “web of trust” — rather than a tree, certificates form an arbitrary graph
- Anyone can sign a certificate
- Most people have more than one signature — I have 65 signatures on my primary PGP key
- Do you know and trust any of my signers?
- See my key at http://www.cs.columbia.edu/~smb/smbpgp.txt
Does the Web of Trust Work?

- Number of signatures alone is meaningless; I can create lots of identities if I want.
- I can even forge names — is the “Angelos Keromytis” who signed my key the same one who’s a professor here? How do you know?
- There are at least six PGP keys purporting to belong to “George W. Bush”. One is signed by “Yes, it’s really Bush!”
- You have to define your own set of trust anchors, as well as policies on how long a signature chain is too long.

- Secure Email
  - PGP and S/MIME
  - Approaches to Protecting Content
  - Certificate Style
- Web of Trust
  - Does the Web of Trust Work?
  - Finding Public Keys
  - Which Style is Better?
- Spam
- Phishing
Finding Public Keys

- Many mailers cache received certificates
- Some organizations list people’s certificates in an LDAP database
- Some people have them on their web site
- For PGP, there are public key servers — anyone can upload keys
- Is that safe? Sure — the security of a certificate derives from the signature, not from where you found it
Which Style is Better?

- PGP was easier to start — it doesn’t need an infrastructure
- Many security and network conferences have “PGP key-signing parties”
- S/MIME is better for official use — it makes it clearer when someone is speaking in an organizational role, since the organization issued the certificate.
- Both have usability issues, though PGP is probably worse
Secure Email
PGP and S/MIME

Spam
Spam Originating Machines
Effective Defenses
Today's Defenses
Blacklisting
Port 25 Blocks
Origin Authentication
SPF Records
DKIM Authentication
The Real Issue with Origin Authentication
Semantic and Keyword Filters
Charging for Email
Phishing
Spam

- We all know what it is...
- Defending against it is very hard
- It is unlikely that the problem will ever go away
Originating Machines

- Originally from the spammer’s own machines — those were blacklisted
- Next: open relays — those have mostly been closed down
- Now: hacked home machines
- Occasionally: routing attacks to hide source
Effective Defenses

Secure Email

PGP and S/MIME

Spam

Spam Originating Machines

Effective Defenses

Today's Defenses
Blacklisting
Port 25 Blocks
Origin Authentication
SPF Records
DKIM Authentication
The Real Issue with Origin Authentication
Semantic and Keyword Filters
Charging for Email

Phishing
Today’s Defenses

- Blacklisting
- Especially blacklisting of “non-mail” machines: dial-ups, home machines, etc.
- Port 25 blocks
- Origin authentication: digital signatures, SPF, DKIM
- Semantic and keyword filters
Blacklisting

- Mostly works, but...
- False positives
- Often, lack of responsiveness by blacklist sites
- Some are trying to dodge lawsuits by spammers
- Others are trying dodge denial-of-service attacks...
- Affects legitimate but unusual users — home users who run their own MTA, some travelers, etc.
Port 25 Blocks

- Many ISPs block outbound port 25
- Force all email to go through ISP’s servers
- Monitor for “too much”
- Demand password (but malware steals passwords anyway)
Origin Authentication

- Concept: prevent spam from forged addresses
- But — most spam isn’t “joe job” spam
- Causes problems with mailing lists
- Causes problems for portable addresses
- SPF — *not* a standard — especially bad in this respect
- Origin authentication better used for whitelists
Columbia’s SPF record (in the DNS):
v=spf1 ip4:128.59.28.0/24
  ip4:128.59.29.0/24  ip4:128.59.59.0/24
  ip4:128.59.62.0/24
  ip4:128.59.28.160/27
  ip4:128.59.29.0/28  ip4:128.59.31.0/24
  ip4:128.59.39.0/24  ~all

Those IP addresses, and no others, are allowed to send mail claiming to be from @columbia.edu addresses.

What if you use your Gmail account with a CU return address?
DKIM Authentication

- Digital signature of (some) mail headers and message body
- Being standardized by the IETF
- Generally done at the originating gateway
- Granularity is generally per-site, but per-user keys are supported (e.g., for laptops for road warriors)
- Public keys are in the DNS, rather than in separate certificates
- Doesn’t change the mail body the way that S/MIME does
- Intended to be lighter-weight
Most people want to permit email from unknown parties
Knowing that the message really is from SomeoneNew@Somewhere.In.the.World doesn’t tell me if it’s spam or not
It prevents “joe jobs”, and it’s good for whitelisting
It doesn’t block spam
We’re seeing the difference between authentication and authorization
Semantic and Keyword Filters

- Look for keywords, improbable text, etc.
- But — spammers include real text excerpts
- Some spam is in attachments, especially image attachments
- Other spam changes words slightly: Viagra → Vِ1agra or V*i*a*g*r*a
- Who has a better polymorphic engine?
Some people suggesting charging for email — email “postage”

Goal: increase the cost to the spammer

Lots of reasons it doesn’t work well — the main one is that the spammers are using hacked machines to send their email.

Requiring postage would mean they’d steal money, too...
Phishing

What is Phishing?
A Phish
What’s Wrong?
The Login Box
The URL Bar
They Want Data...
Some Mail Headers
Other Issues
Tricks with URLs
Final Thoughts on Phishing
What is Phishing?

- Spoofed emails, purportedly from a financial institution
- Ask you to login to “reset” or “revalidate” your account
- Often claim that your account has been suspended
From: no-reply@flagstarbanking2.com
To: undisclosed-recipients:;
Subject: YOUR ACCOUNT HAS BEEN SUSPENDED !!!
Date: Fri, 29 Sep 2006 09:29:25 -0500

... 

If you fail to provide information about your account you’ll discover that your account has been automatically deleted from Flagstar Bank database.

Please click on the link below to start the update process:

https://www.flagstar.com/Signon.cgi?update
Flagstar Bank
What’s Wrong?

- The URL is a booby trap:
  
  **Fake URL warning**
  
  
  Open it anyway?

- When I clicked on it, I was actually redirected to a site in Colombia, via yet another indirection...

- The login page appears identical to the real one

- (One of the web sites I visited seemed to have several variant “bank” pages)
The Login Box

Welcome to Flagstar Bank's Internet Banking

Registered Users, Please Enter Your User ID and Password. First time users, please click here to register.

Forgot your Internet Banking Password? Click here to reset it yourself - OR - Click here to have Flagstar Bank reset it for you.

User ID:

Password:

Login
The URL Bar

Welcome to Flagstar Bank's Internet Banking

Registered Users, Please Enter Your User ID and Password. First time users, please click here to register.

Forgot your Internet Banking Password? Click here to reset it yourself - OR - Click here to have Flagstar Bank reset it for you.

User ID:
Password:
Login
Please complete the fields below to recover account.

Required fields are in red.

First Name
Last Name
Card Number
Expiration Date
Electronic Signature (ATM PIN)
Social Security Number (SSN)
Home Phone #
Email Address

☐ Click here if you want to receive confirmation email.
☐ Click here if you do not want to receive confirmation email.

Note: You will receive the confirmation email within 48 hours.
Secure Email
PGP and S/MIME
Spam
Phishing
What is Phishing?
A Phish
What’s Wrong?
The Login Box
The URL Bar
They Want Data...

Some Mail Headers

Received: from plesk.salesforcefoundation.org
       ([198.87.81.9])
       by cs.columbia.edu (8.12.10/8.12.10)
       (version=TLSv1/SSLv3 cipher=DHE-RSA-AES256-SHA
        bits=256 verify=NOT) for <smb@cs.columbia.edu>

Received: from adsl-68-20-44-198.dsl.chcgil.ameritech.net
       (68.20.44.198) by 198.87.81.11

Where does plesk.salesforcefoundation.org come from? It is asserted by the far side. The 198.87.81.9 is derived from the IP header, and is hard to forge (but stay tuned for routing attacks, in a few weeks). A DNS lookup on 198.87.81.9 isn’t very helpful; the mapping is controlled by the address owner, not the name owner.
Other Issues

- Why is the email from flagstarbanking2.com?
- The domain for the bank is flagstar.com — no “ing” and no “2”.
- *That’s legit!* — the real web site for their online service is flagstarbanking2.com
- We have trained users to accept weird, seemingly gratuitous differences; it can make life easier for the phisher
Tricks with URLs

- http://cnn.com@some.other.site/foocnn.com is a userid
- http://2151288839/foo2151288839 is 128.58.16.7,cluster.cs.columbia.edu
- http://rds.yahoo.com/_ylt=A0g...http%3a/So the search engine knows what you clicked on
Final Thoughts on Phishing

- We have the basic technical mechanisms to authenticate email and web sites
- Human interaction with these mechanisms remains a very challenging problem
- Security is a *systems problem*