COLLEGE OF SOCIAL & BEHAVIORAL SCIENCES
School of Information

Communications Privacy II: Public Key Cryptography Background + RSA

Information Privacy with Applications

David Sidi (dsidi@email.arizona.edu)

Warm-up

- Explain the difference between symmetric encryption and asymmetric encryption, from POTL

Small mention of interesting things

- New onion addresses are in alpha
- Assignment 2 is out
- Demo: setting up a hidden service (I forgot to do this last time)
- Demo: connecting to the OP on a control port
- Stem

COLLEGE OF SOCIAL & BEHAVIORAL SCIENCES
School of Information

Continuing last time: Communication Privacy

Cryptography is useful when it is difficult to secure a channel

- Confidentiality in FF voice communication requires that no unwanted third party is listening in
- However hard that is, phone communication presupposes it too. In addition, it requires that the call isn't intercepted while in transit
- Interception: Face to face < Copper wire < Radio link < Optical link (**POTL** 11)

Is privacy easier to achieve if privacy failure is easier to detect?

- Is it harder to read a letter surreptitiously over someone's shoulder than to listen to a conversation surreptitiously?
  - 2 minutes then rejoin

Is privacy is easier to achieve if violating privacy is easier to detect?

- If attackers go for the stealthiest option, the greater threat to letter communication is interception of a letter in transit
- One way to go: secure the channel: US postal service. Remember the history there in colonial America?
- Another way: encrypt the communications
  - notice you still have a part of the channel to secure (think back to FF case, and our “Layer 8+” discussion)

Tamper detection in electronic communication is hard

- Envelopes (weakly) detect tampering in written communication
- there is no analog for encrypted communications
  - can check authenticity and integrity, though
  - also, see QKD

Tamper proof key distribution + one time pads

- OTP create *key management* problems, as we'll see. QKD helps with that
- "A hub-and-spoke network has been operated by Los Alamos National Laboratory since 2011. All messages are routed via the hub. The system equips each node in the network with quantum transmitters—i.e., lasers—but not with expensive and bulky photon detectors. Only the hub receives quantum messages. To communicate, each node sends a one-time pad to the hub, which it then uses to communicate securely over a classical link. The hub can route this message to another node using another one time pad from the second node." (link)

One-time pad systems (OTP) are information-theoretically secure, subject to some conditions

- Key is as long as the message
- Key is random
- Key is secret
- Key is not reused
  - Creates a key distribution problem

OTP systems create 101 key-management problems

- Encode the time and place of an event as 8 two-digit decimal numbers
- YYYY MM DD hh mm NS EW, where Y:=year, M:= month, D:=day, h:=hour, m:=minute, NS:=north-south street number, EW:=east-west street number
- Say the message is

  19 99 12 30 15 25 01 44

OTP systems create 101 key-management problems

- Say the mesage is

  19 99 12 30 15 25 01 44
- Key is a random set of 8 two-digit numbers

  64 25 83 09 76 23 55 72
- Add the key to the message, “forgetting any carrying” (i.e. add in $\mathbb{Z}_{10}$):

  19 99 12 30 15 25 01 44

  64 25 83 09 76 23 55 72

  ---

  73 14 95 39 81 48 56 16

OTP systems create 101 key-management problems

- No one without the key can decrypt the message, there isn't enough information for a ciphertext-only attack
- Suppose the message and the key were, respectively,

  20 00 01 11 10 45 05 23, and

  53 14 94 28 71 03 51 93,

  then the ciphertext would be the same

OTP systems create 101 key-management problems

- OTP protects against ciphertext-only attacks; known plaintext attacks are another story
- Say the event you're encoding happens on 12/30/1999 at 3:25 on the corner of 1st and 44th, and Eve has the ciphertext
- She subtracts to get the key, "forgetting borrowing" (i.e. subtract in $\mathbb{Z}_{10}$):

  73 14 95 39 81 48 56 16

  19 99 12 30 15 25 01 44

  ---

  64 25 83 09 76 23 55 72

OTP systems create 101 key-management problems

- Lesson: OTP is only as secure as the key management protocols that go with it
- This can be an organizational nightmare
  - Leave it to the Soviets to use...central planning

For reasons that are still unclear, a serious mistake was made in the early months of 1942. Rather than making exactly two copies of the key sheets, they made four. These excess keys then entered the inventory and remained in use for several years. Western intelligence noted and exploited the multiple use of the keys, with disastrous results for Soviet security. Under the code name Venona, cryptanalytic study of the reused “one-time” keys went on for decades. The system was used for the most sensitive Soviet information, and the Americans and the British studied it in hopes of identifying Soviet “moles” thought to be operating at the highest levels of their intelligence establishments. (POTL 19)

OTP systems create 101 key-management problems

- Lesson: OTP is only as secure as the key management protocols that go with it
- This can be an organizational nightmare
  - Leave it to the Soviets to use...central planning
- (Enter QKP)

Different cryptographic systems have different strengths

- “A cryptosystem is considered secure when an opponent cannot break it under reasonable circumstances, in a reasonable amount of time, at a reasonable cost. The term “reasonable” is perforce vague.” (POTL 26)

“Reasonable Circumstances”

- Ciphertext only
- Known plaintext: attacker can observe a plaintext and its encryption
- Chosen plaintext: attacker picks the plaintext to be encrypted
- Chosen ciphertext: attacker picks the ciphertext to be decrypted
  - non-malleability: the attacker cannot change the ciphertext so that the correspondnig plaintext is changed in a controlled way

“Reasonable Time”

- “Workfactor:” number of operations required to break a cryptographic system
- What ‘operations’ means depends: they may not be elementary computer instructions
  - For example, if searching space of keys: operations are encryptions, which could be several hundred instructions

Workfactors and their significance

- Assume perfectly parallel problems
- 2^{30}: trivial (minutes) by one computer at 1 GhZ
- 2^{60}: 11 to 12 days (with 2^{20} processors in parallel at 2^{20} instructions per second)
- 2^{90}: 30 years (with 2^{30} processors in parallel at 2^{30} instructions per second)
- 2^{120}: 30,000 years (with 2^{60} processors in parallel at 2^{60} instructions per second)

Estimating workfactor significance: what could go wrong?

- RSA challenge: in 1977, it was estimated that the time taken to factor the 426 bit number would be **4×10^{16} years**
- n = 114381625757888676692357799761466120102182967212423625625618429357069352457338978305971235639587050589890751475992900268795435 41
- This is from Martin Gardner's Scientific American article, and came to be known as RSA-129
- Solved in 1994 (~ **17 years** later)

“Reasonable time” is relative

- How long is too long for decryption?
  - The Venona messages were studied for nearly 40 years in hopes that they would reveal the identities of spies who had been young men in the 1930s and who might have been the senior intelligence officers of the 1970s
- Sometimes keys are ephemeral, so are only helpful for a small window of messages going forward
  - What’s an example of an ephemeral key from Tor?

“Reasonable cost”

- Example: NSA’s Utah Data Center
- The planned structure provides 1 to 1.5 million square feet (90,000–140,000 m2), with 100,000 square feet (9,000 m2) of data center space and more than 900,000 square feet (84,000 m2) of technical support and administrative space. It is projected to cost $1.5–2 billion. A report suggested that it will cost another $2 billion for hardware, software, and maintenance.

- “cryptography can best be thought of as a mechanism for extending the confidentiality and authenticity of one piece of information (the key) to another (the message).” (POTL 34)

Key compromise means different things depending on the key's use

- Authentication keys can be revoked, and no authentications will still go through with those keys
- Keys used for privacy can also be revoked, but all messages ever sent with a key must be regarded as compromised
- A revoked key can be used in the future for old encryptions, but there is no corresponding notion of "old authentications"

"Cryptography is the only technique capable of providing security to messages transmitted over channels entirely out of the control of either the sender or the receiver." (POTL 35)

Rivest's riposte

- Diffie missed something: Ron Rivest's idea of chaffing and winnowing for confidentiality
- OK, still cryptography, but not encryption, so an important qualification to Diffie's comment
- Not encryption!
- Actually, he missed two things: What is another example of an approach to confidentiality that does not use encryption?

Rivest's riposte

- Chaffing and Winnowing (C&W) arose amid the same concerns about key escrow, clipper chips, etc. that POTL had in mind
- Key idea: Uses obfuscation to achieve confidentiality over an insecure channel
- A kind of compulsion resistance for cryptography development!

Chaffing and Winnowing is about adding and removing noise

- Sending a message has two parts
  - authenticating (adding MACs)
  - adding “chaff”
- Receiving a message requires removing the “chaff”

Chaff is a set of fake packets

- Chaff packets are not part of the real message
- The MAC of chaff doesn't check, so intended recipients can discard them

(1,Hi Larry,532105)
(1,Hi Bob,465231)
(2,Meet me at,782290)
(2,I'll call you at,793122)
(3,6PM,891231)
(3,7PM,344287)
(4,Yours-Susan,553419)
(4,Love-Alice,312265)

Senders append MACs

- Message is broken into packets by the sender
- MACs are appended to each packet (note: packet is still in the clear)
- MAC is a function of a hash of the message contents, and a shared authentication key
- a serial number can also be added

Chaffing and Winnowing

- Confidentiality of C&W depends on the MAC algorithm, on how the original message is broken into packets, and on how the chaffing is done
- MAC should be indistinguishable from a random function

Public Key Cryptography

- Key idea: Encryption key is public, decryption key is private
- Question: The public key can also be used for decryption, and the private one for encryption. When?

COLLEGE OF SOCIAL & BEHAVIORAL SCIENCES
School of Information

The Rivest-Shamir-Adleman (RSA) Cryptosystem

RSA requires a modulus that is the product of two primes

- randomly choose a large integer n = pq, called the RSA modulus, with p and q prime
- take the group $(\mathbb{Z}/n\mathbb{Z})^*$
- p and q of almost equal length
- There are factoring algorithms that do better with p or q of a special form, but there are only a few instances of that form. With a cryptographic pseudo-random number generator (CPRNG), the probability of getting one is negligible

RSA requires an encryption exponent

- $\varphi(n) = (p - 1)(q - 1)$ is Euler's phi (this is the order of $(\mathbb{Z}/n\mathbb{Z})^*$)
- choose an encryption exponent e such that
  - $1 <= e <= \varphi(n)$
  - e coprime with $\varphi(n)$: $\gcd(e, \varphi(n)) = 1$

RSA requires a decryption exponent

- compute a decryption exponent d
- 1 <= d <= φ(n)
- ed ≡ 1 (mod φ(n))
  - found with extended euclidean algorithm, since gcd(e, φ(n)) = 1

RSA encrypts messages encoded as integers

- message is an integer m with $0 < m < n$
  - can encode $m_1 m_2 \cdots m_k$ as such an m; a block version of RSA
- encryption of a message m is $m^e \pmod n$; decryption is $(m^e)^d \pmod n$
- we need to show that $(m^e)^d = m$

COLLEGE OF SOCIAL & BEHAVIORAL SCIENCES
School of Information

RSA relies on the difficulty of prime factorization

- choose a large number $n = pq$, with p,q prime
- $\varphi(n)$ is Euler's phi
- choose exponents e,d such that
  - $1 <= e,d <= \varphi(n)$
  - e coprime with $\varphi(n)$
  - $ed \equiv 1 \pmod{\varphi(n)}$
- message is an integer m with $0 < m < n$
- encryption of m is

  $m^e \pmod n$
- decryption is

  $(m^e)^d \pmod n$

RSA is a cryptosystem

- To show that RSA is a cryptosystem, we need to show that the encryption operation is invertible

RSA is a cryptosystem

- $ed \equiv 1 \pmod{\varphi(n)}$ implies $ed = 1 + \ell\varphi(n)$
- $(m^e)^d = (m^{1+\ell\varphi(n)}) = m(m^{\ell\varphi(n)}) = m(m^{\ell(p-1)(q-1)}) = m(m^{(p-1)})^{\ell(q-1)}$
- If $p \mid m$, $(m^e)^d \equiv m \pmod{p}$ is trivial. (why?)
- Otherwise, by Fermat's little theorem, $m^{(p-1)} \equiv 1 \pmod{p}$, so $m(m^{(p-1)})^{\ell(q-1)} \equiv m \pmod{p}$
- The case is exactly similar for q, so we have $(m^e)^d \equiv m \pmod{pq}$
- $0 < m < pq$, so it is established that $(m^e)^d = m$

- Lots of the details you've just seen about the usual implementation of RSA are inessential
- The core idea of RSA is from group theory: exponentiation by an element coprime with the group order is an invertible automorphism
  - Klaus Lux occasionally teaches a cryptography course here; take it to learn more

RSA is a partially-homomorphic system

- What is homomorphic encryption?
- $(gh)^e = g^e h^e$, given that we're in an abelian group, so RSA can be used for partially homomorphic encryption

Public Key Cryptography

- Encryption key is public, decryption key is private
- we will discuss ‘New Directions’ next time as well
- Key management is centralized only for the public key; private keys are kept secret by the individual key holders
- But private key must be securely generated, and kept safe
  - mitigation is different for non-interactive and interactive communications, however