Message Integrity, Cryptographic Hash Functions and Digital Signatures: Part 2

Gaurav S. Kasbekar

Dept. of Electrical Engineering

IIT Bombay

References

- J. Kurose, K. Ross, "*Computer Networking: A Top Down Approach*", Sixth Edition, Pearson Education, 2013
- C. Kaufman, R. Perlman, M. Speciner, "*Network Security: Private Communication in a Public World*", Pearson Education, 2nd edition, 2002
- A. Tanenbaum, D. Wetherall, "Computer Networks", Fifth Edition, Pearson Education, 2012

Some Properties of Cryptographic Hash Functions

Simple Model for Crytographic Hash Function Output

- A cryptographic hash function should appear to be a completely random mapping from input to output
  - similar to a symmetric-key encryption algorithm
- For a crypt. hash function that produces an l bit output, "completely random" means the following:
  - for a given input, output selected from among the 2^l possible outputs uniformly at random
  - for a different input, output *independently* selected uniformly at random from the 2^l possible outputs
  - and so on
- Reason for requiring mapping to appear completely random:
  - otherwise it may be possible to efficiently find an input that results in a pre-specified output or to find two different inputs that result in the same output
- Hence, a simple model for a crypt. hash function with a l bit output is to assume that:
  - for any given input, the output is selected uniformly at random from the 2^l possible outputs

What Should the Number of Output Bits l Be?

- Recall: for achieving message integrity:
  - it should be computationally infeasible to find a message m that has a pre-specified hash value h, i.e., m such that $H(m) = h$
- Suppose intruder finds the hash values of different random inputs until he/ she finds an input that has hash value h
- How many inputs must intruder try, to find an input with hash value h with high probability?
- If intruder tries n different inputs, expected number of inputs whose hash value equals h is:
  - $\frac{n}{2^l}$
  - **Proof**:
    - Let $X_i = \begin{cases} 1, & \text{if } i'\text{th input has hash value } h, \\ 0, & \text{else.} \end{cases}$
    - Let $Y = X_1 + \cdots + X_n$
    - Then $E(Y) = nE(X_1) = \frac{n}{2^l}$
- So number of inputs that must be tried, on average, until success:
  - $n = 2^l$
- Hence, larger the value of l, more difficult it is for intruder
- E.g.:
  - if $l = 64$, then intruder must try $2^{64} \approx 1.8 \times 10^{19}$ inputs, which is computationally infeasible
  - if $l = 32$, then intruder must try $2^{32} \approx 4.3 \times 10^{9}$ inputs, which is feasible
- Above analysis suggests that a value of $l = 64$ is adequate for achieving message integrity

What Should the Number of Output Bits l Be? (contd.)

- Recall: above analysis suggests that a value of $l = 64$ is adequate for achieving message integrity
- However, for some applications, we require the following stronger property:
  - it is computationally infeasible to find two messages m and m' such that $H(m') = H(m)$
- Later, we will discuss an example of such an application
- Suppose we find the hash values of n inputs
- Hash value of each input selected uniformly at random from the 2^l possible outputs
- Let $p = P$(two different inputs m and m' have the same hash value, i.e., $H(m') = H(m)$)
- How large must n be such that $p \geq \frac{1}{2}$?
- **Exercise**: For n larger than $\approx 2^{l/2}$, $p \geq \frac{1}{2}$
- That is, *if we find the hash values of $n \approx 2^{l/2}$ different inputs, then with a high probability $\left(p \geq \frac{1}{2}\right)$, we will find two inputs with the same hash value*

What Should the Number of Output Bits l Be? (contd.)

- Recall: if we find the hash values of $n \approx 2^{l/2}$ different inputs, then with a high probability $\left(p \geq \frac{1}{2}\right)$, we will find two inputs with the same hash value
- E.g.:
  - ❑ if $l = 64$, then intruder must try $2^{32} \approx 4.3 \times 10^9$ inputs, which is feasible
  - ❑ if $l = 128$, then intruder must try $2^{64} \approx 1.8 \times 10^{19}$ inputs, which is computationally infeasible
- So in applications where we require the property:
  - it is computationally infeasible to find two messages m and m' such that $H(m') = H(m)$,

  a value of $l = 64$ is *not* adequate
- Hence, a value of $l = 128$ or larger is used

What Should the Number of Output Bits l Be? (contd.)

- Recall: $p = P(\text{out of } n \text{ inputs, two different inputs } m \text{ and } m' \text{ have the same hash value, i.e., } H(m') = H(m))$
- p can be found as in the solution to "*The Birthday Problem*":
  - ❑Out of a set of n randomly chosen people, what is the probability that two of them have the same birthday?
- Hence, an attack on a cryptographic hash function, in which intruder tries n inputs to find two different inputs with the same hash value is called the "*The Birthday Attack*"

Example of The Birthday Attack

- Tom applies for a tenured faculty position
- Requests his department chairperson, Marilyn, who thinks highly of his work, for a letter of recommendation
- Marilyn outlines the contents of the letter to her secretary, Ellen, and asks her to compose the letter
  - After Marilyn's approval, Ellen will send the letter itself to the Dean and Marilyn will compute and email the 64 −bit hash value of the letter to the Dean for verification
- However, Ellen has a grudge against Tom and wants to damage his application
- So she creates two sets of 2^{32} letters each, one set providing a positive recommendation and the other set providing a negative recommendation

Example of The Birthday Attack (contd.)

- Set of positive letters:

Dear Dean Smith,

This [letter | message] is to give my [honest | frank] opinion of Prof. Tom Wilson, who is [a candidate | up] for tenure [now | this year]. I have [known | worked with] Prof. Wilson for [about | almost] six years. He is an [outstanding | excellent] researcher of great [talent | ability] known [worldwide | internationally] for his [brilliant | creative] insights into [many | a wide variety of] [difficult | challenging] problems.

He is also a [highly | greatly] [respected | admired] [teacher | educator]. His students give his [classes | courses] [rave | spectacular] reviews. He is [our | the Department's] [most popular | best-loved] [teacher | instructor].

[In addition | Additionally] Prof. Wilson is a [gifted | effective] fund raiser. His [grants | contracts] have brought a [large | substantial] amount of money into [the | our] Department. [This money has | These funds have] [enabled | permitted] us to [pursue | carry out] many [special | important] programs, [such as | for example] your State 2000 program. Without these funds we would [be unable | not be able] to continue this program, which is so [important | essential] to both of us. I strongly urge you to grant him tenure.

Example of The Birthday Attack (contd.)

- Set of negative letters:

Dear Dean Smith,

This [letter | message] is to give my [honest | frank] opinion of Prof. Tom Wilson, who is [a candidate | up] for tenure [now | this year]. I have [known | worked with] Prof. Wilson for [about | almost] six years. He is a [poor | weak] researcher not well known in his [field | area]. His research [hardly ever | rarely] shows [insight in | understanding of] the [key | major] problems of [the | our] day.

Furthermore, he is not a [respected | admired] [teacher | educator]. His students give his [classes | courses] [poor | bad] reviews. He is [our | the Department's] least popular [teacher | instructor], known [mostly | primarily] within [the| our] Department for his [tendency | propensity] to [ridicule | embarrass] students [foolish | imprudent] enough to ask questions in his classes.

[In addition | Additionally] Tom is a [poor | marginal] fund raiser. His [grants | contracts] have brought only a [meager | insignificant] amount of money into [the | our] Department. Unless new [money is | funds are] quickly located, we may have to cancel some essential programs, such as your State 2000 program. Unfortunately, under these [conditions | circumstances] I cannot in good [conscience | faith] recommend him to you for [tenure | a permanent position].

Example of The Birthday Attack (contd.)

- Recall: Ellen creates two sets of 2^{32} letters each, one set providing a positive recommendation and the other set providing a negative recommendation
- With a high probability, she finds one positive letter m and one negative letter m' with the same hash value, i.e., $H(m) = H(m')$
- Ellen emails the positive letter, m, to Marilyn for approval and the negative letter, m', to the Dean
- Hash value verification by Dean succeeds since $H(m) = H(m')$
- **Note**: P(Ellen finds a positive letter and a negative letter with the same hash value):
  - ❑ $\approx \frac{1}{4}$