IF YOU DID NOT GET EMAIL FROM ME

IF you did not get email from me then see me NOW so I can put you on the list.
I mean RIGHT NOW!!!!!!!!!!!!!!!
If I said see me after class
you might forget. This has actually happened.

Something Wrong With All Ciphers So Far

September 12, 2019

Eve CAN tell...

Let C be any of Shift, Affine, Gen Sub, Vig, Matrix, Playfair, Rail
(NOT one-time pad, Book-Vig, Autokey-Vig)
Assume Eve does not know how to crack C.
But: Eve can still tell if two messages are the same or not.
EASILY!
Is this a problem?

Eve CAN tell...

Let C be any of Shift, Affine, Gen Sub, Vig, Matrix, Playfair, Rail (NOT one-time pad, Book-Vig, Autokey-Vig)
Assume Eve does not know how to crack C.
But: Eve can still tell if two messages are the same or not.
EASILY!
Is this a problem?

YES! Eve knows that the message will say where the spy is. The message Will be of the form *city,state* (without punctuation).

Eve CAN tell. . .

Let C be any of Shift, Affine, Gen Sub, Vig, Matrix, Playfair, Rail (NOT one-time pad, Book-Vig, Autokey-Vig)
Assume Eve does not know how to crack C.
But: Eve can still tell if two messages are the same or not.
EASILY!
Is this a problem?

YES! Eve knows that the message will say where the spy is. The message Will be of the form *city,state* (without punctuation).
Alice sends to Bob adecn aapad ecnaa pxuaq.

Eve CAN tell...

Let C be any of Shift, Affine, Gen Sub, Vig, Matrix, Playfair, Rail (NOT one-time pad, Book-Vig, Autokey-Vig)
Assume Eve does not know how to crack C.
But: Eve can still tell if two messages are the same or not.
EASILY!
Is this a problem?

YES! Eve knows that the message will say where the spy is. The message Will be of the form *city,state* (without punctuation).
Alice sends to Bob adecn aapad ecnaa pxuaq.

Eve notices adecnaap adecnaap xuaq.

Eve CAN tell. . .

Let C be any of Shift, Affine, Gen Sub, Vig, Matrix, Playfair, Rail (NOT one-time pad, Book-Vig, Autokey-Vig)
Assume Eve does not know how to crack C.
But: Eve can still tell if two messages are the same or not.
EASILY!
Is this a problem?

YES! Eve knows that the message will say where the spy is. The message Will be of the form *city,state* (without punctuation).
Alice sends to Bob adecn aapad ecnaa pxuaq.

Eve notices adecnaap adecnaap xuaq.

Eve knows that the city and state are the same!

What Does Eve Know?

Cities with a state's name. * means no longer a city.

What Does Eve Know?

Cities with a state's name. * means no longer a city.

Alabama*, Arizona*, Arkansas, California, Colorado*, Delaware, Florida, New Georgia*, Idaho, Illinois*, Indianapolis, Iowa, Jersey, Kansas, Maryland*, Minneapolis, Minnesota, Mississippi*, Missouri, Montana, Nebraska, Nevada*, New York, Ohio, Oklahoma, Oregon, Tennessee*, Texas, Utah*, Virginia*, Virginia Beach, Wisconsin Dells, Wisconsin Rapids.

What Does Eve Know?

Cities with a state's name. * means no longer a city.

Alabama*, Arizona*, Arkansas, California, Colorado*, Delaware, Florida, New Georgia*, Idaho, Illinois*, Indianapolis, Iowa, Jersey, Kansas, Maryland*, Minneapolis, Minnesota, Mississippi*, Missouri, Montana, Nebraska, Nevada*, New York, Ohio, Oklahoma, Oregon, Tennessee*, Texas, Utah*, Virginia*, Virginia Beach, Wisconsin Dells, Wisconsin Rapids.

There are 33 such cities, 22 of which still exist.
Eve's search for the spy is reduced!

How to Fix This?

Problem: If C is any of the ciphers discussed (except 1-time pad, Book-Vig, Autokey-Vig) then Eve can tell when two messages are the same.

Discuss: Is there a cipher for which Eve cannot tell this?

How to Fix This?

Problem: If C is any of the ciphers discussed (except 1-time pad, Book-Vig, Autokey-Vig) then Eve can tell when two messages are the same.

Discuss: Is there a cipher for which Eve cannot tell this?
Need that even if $x = y$ could have $C(x) \neq C(y)$.

Discuss: How can we do that?

How to Fix This?

Problem: If C is any of the ciphers discussed (except 1-time pad, Book-Vig, Autokey-Vig) then Eve can tell when two messages are the same.

Discuss: Is there a cipher for which Eve cannot tell this?
Need that even if $x = y$ could have $C(x) \neq C(y)$.

Discuss: How can we do that?

Use a very long key and keep using different parts of it, which is the 1-time pad, Book-Vig, Autokey-Vig. Is there an easier way?

How to Fix This?

Problem: If C is any of the ciphers discussed (except 1-time pad, Book-Vig, Autokey-Vig) then Eve can tell when two messages are the same.

Discuss: Is there a cipher for which Eve cannot tell this?
Need that even if $x = y$ could have $C(x) \neq C(y)$.

Discuss: How can we do that?

Use a very long key and keep using different parts of it, which is the 1-time pad, Book-Vig, Autokey-Vig. Is there an easier way?

Discuss: Can we do this without a long key?

How to Fix This Without a Long Key

Obstacle: All of our ciphers are deterministic. Need Rand.

How to Fix This Without a Long Key

Obstacle: All of our ciphers are deterministic. Need Rand.
Recall Deterministic Shift: Key is $s \in S$.

1. To send message $(m_1, \ldots, m_L)$ send $(m_1 + s, \ldots, m_L + s)$
2. To decode message $(c_1, \ldots, c_L)$ find $(c_1 - s, \ldots, c_L - s)$

How to Fix This Without a Long Key

Obstacle: All of our ciphers are deterministic. Need Rand.
Recall Deterministic Shift: Key is $s \in S$.

1. To send message $(m_1, \ldots, m_L)$ send $(m_1 + s, \ldots, m_L + s)$
2. To decode message $(c_1, \ldots, c_L)$ find $(c_1 - s, \ldots, c_L - s)$

Randomized shift: Key is a function $f : S \to S$.

1. To send message $(m_1, \ldots, m_L)$ (each m_i is a character)
   1.1 Pick random $r_1, \ldots, r_L \in S$.
   1.2 Send $((r_1; m_1 + f(r_1)), \ldots, (r_L; m_L + f(r_L)))$
2. To decode message $((r_1; c_1), \ldots, (r_L; c_L))$
   2.1 Find $(c_1 - f(r_1), \ldots, c_L - f(r_L))$

Example

The key is $f(r) = 2r + 7$. Alice wants to send NY,NY which we interpret as nyny.
Need four shifts.

Pick random $r = 4$, so first shift is $2 \times 4 + 7 = 15$
Pick random $r = 10$, so second shift is $2 \times 10 + 7 = 1$
Pick random $r = 1$, so third shift is $2 \times 1 + 7 = 9$
Pick random $r = 17$, so fourth shift is $2 \times 17 + 7 = 15$

Send (4;C), (10;Z), (1;W), (17;N)

Eve will not be able to tell that is of the form XYXY.

PROS and CONS of Randomized Shift

Discuss

PROS and CONS of Randomized Shift

Discuss

PRO: If Alice sends NY,NY Eve can't tell its XYXY.

PROS and CONS of Randomized Shift

Discuss

PRO: If Alice sends NY,NY Eve can't tell its XYXY.

PRO: More generally, Eve cannot tell if two messages are the same.

PROS and CONS of Randomized Shift

Discuss

PRO: If Alice sends NY,NY Eve can't tell its XYXY.

PRO: More generally, Eve cannot tell if two messages are the same.

CON: More effort on Alice and Bob's part.

PROS and CONS of Randomized Shift

Discuss

PRO: If Alice sends NY,NY Eve can't tell its XYXY.

PRO: More generally, Eve cannot tell if two messages are the same.

CON: More effort on Alice and Bob's part.

Question: Is Randomized Shift crackable? Discuss.

Long Aside: The Birthday Paradox

September 12, 2019

Birthday Paradox

Let $m < n$. We figure out m, n later.
We will put m balls into n boxes uniformly at random.
Goal: How big does m have to be before the prob that some box has 2 balls is $\geq \frac{1}{2}$?

Birthday Paradox

Let $m < n$. We figure out m, n later.
We will put m balls into n boxes uniformly at random.
Goal: How big does m have to be before the prob that some box has 2 balls is $\geq \frac{1}{2}$?

We ask opp: What is prob that NO box has ≥ 2 balls?

- Number of ways to put balls into boxes: n^m
- Number of ways to put balls into boxes: so that no box has ≥ 2 balls: $n(n-1)\cdots(n-m+1)$

The probability is

$$\frac{n(n-1)(n-2)\cdots(n-m+1)}{n^m}$$

Approx

$$\frac{n(n-1)(n-2)\cdots(n-m+1)}{n^m}$$

$$= \frac{n}{n} \times \frac{n-1}{n} \times \frac{n-2}{n} \times \cdots \times \frac{n-m+1}{n}$$

$$= 1 \times \left(1 - \frac{1}{n}\right) \times \left(1 - \frac{2}{n}\right) \times \cdots \times \left(1 - \frac{m-1}{n}\right)$$

Approx

$$\frac{n(n-1)(n-2)\cdots(n-m+1)}{n^m}$$

$$= \frac{n}{n} \times \frac{n-1}{n} \times \frac{n-2}{n} \times \cdots \times \frac{n-m+1}{n}$$

$$= 1 \times \left(1 - \frac{1}{n}\right) \times \left(1 - \frac{2}{n}\right) \times \cdots \times \left(1 - \frac{m-1}{n}\right)$$

Recall: $e^{-x} \sim 1 - x$ for x small. So we have

$$\sim e^{-1/n} \times e^{-2/n} \times \cdots e^{-(m-1)/n} = e^{-(1/n)(1+2+\cdots+(m-1))}$$

$$\sim e^{-m^2/2n}$$

Real Numbers!

If $m < n$ and you put m balls in n boxes at random then prob that ≥ 2 balls in same box is approx:

$$1 - e^{-m^2/2n}$$

Recall: Our goal is to find m such that prob of 2 in the same box is $\geq \frac{1}{2}$. Hence we need $1 - e^{-m^2/2n} > \frac{1}{2}$:

$$e^{-m^2/2n} < \frac{1}{2}$$

$$-\frac{m^2}{2n} < \ln(0.5) \sim -0.7$$

$$\frac{m^2}{2n} > 0.7$$

$$m > (1.4n)^{1/2}$$

Real Numbers!

If $m > (1.4n)^{1/2}$ and you put m balls in n boxes at random then prob that ≥ 2 balls in same box is over $\frac{1}{2}$.

$n = 365$.
$m = \lceil (1.4n)^{1/2} \rceil = 23$

Birthday Paradox: If there are 23 people in a room then prob two have the same birthday is $> \frac{1}{2}$.

How We Use: If $\sim n^{1/2}$ balls put into n boxes then prob 2 in same box is large.

Alternative Proof

Prob balls i, j in same box is $\frac{n}{n^2} = \frac{1}{n}$.
Prob balls i, j NOT in same box is $1 - \frac{1}{n}$.

Alternative Proof

Prob balls i, j in same box is $\frac{n}{n^2} = \frac{1}{n}$.
Prob balls i, j NOT in same box is $1 - \frac{1}{n}$.

Prob NO pair is in same box: Want to say $(1 - \frac{1}{n})^{\binom{m}{2}}$.

Alternative Proof

Prob balls i, j in same box is $\frac{n}{n^2} = \frac{1}{n}$.
Prob balls i, j NOT in same box is $1 - \frac{1}{n}$.

Prob NO pair is in same box: Want to say $(1 - \frac{1}{n})^{\binom{m}{2}}$.

Not quite. Would be true if they are all ind. But good approx.

Alternative Proof

Prob balls i, j in same box is $\frac{n}{n^2} = \frac{1}{n}$.
Prob balls i, j NOT in same box is $1 - \frac{1}{n}$.

Prob NO pair is in same box: Want to say $(1 - \frac{1}{n})^{\binom{m}{2}}$.

Not quite. Would be true if they are all ind. But good approx.

Prob NO pair is in same box $< (1 - \frac{1}{n})^{\binom{m}{2}} \sim e^{-m^2/2n}$.
Prob SOME pair is in same box $> 1 - e^{-m^2/2n}$.
Same as before.

Three Balls in a Box

Prob balls i, j, k in same box is $\frac{n}{n^3} = \frac{1}{n^2}$.
Prob balls i, j, k NOT in same box is $1 - \frac{1}{n^2}$.

Three Balls in a Box

Prob balls i, j, k in same box is $\frac{n}{n^3} = \frac{1}{n^2}$.
Prob balls i, j, k NOT in same box is $1 - \frac{1}{n^2}$.

Prob NO triple is in same box: $\sim (1 - \frac{1}{n^2})^{\binom{m}{3}} \sim e^{-m^3/6n^2}$
Prob SOME triple is in same box: $\sim 1 - e^{-m^3/6n^2}$

Real Numbers!

If $m < n$ and you put m balls in n boxes at random then prob that ≥ 3 balls in same box is approx:

$$1 - e^{-m^3/6n^2}$$

Real Numbers!

If $m < n$ and you put m balls in n boxes at random then prob that ≥ 3 balls in same box is approx:

$$1 - e^{-m^3/6n^2}$$

To get this $> \frac{1}{2}$ need $1 - e^{-m^3/6n^2} > \frac{1}{2}$

$$e^{-m^3/6n^2} < \frac{1}{2}$$

$$-\frac{m^3}{6n^2} < \ln(0.5) \sim -0.7$$

$$m > (4.2n)^{2/3}$$

Real Numbers!

If $m > (4.2n)^{2/3}$ and you put m balls in n boxes at random then prob that ≥ 3 balls in same box is over $\frac{1}{2}$.

$n = 365$.
$m = \lceil (4.2n)^{2/3} \rceil = 82$

Birthday Paradox: $n = 365$ then need $m \geq 82$. SO if 82 people in a room prob is $> \frac{1}{2}$ that three have same bday!

How We Use: If $\sim n^{2/3}$ balls put into n boxes then prob 3 in same box is large.

Recap and Generalize

1. $\sim n^{1/2}$ balls put into n boxes, prob 2 in same box.
2. $\sim n^{2/3}$ balls put into n boxes, prob 3 in same box.
3. $\sim n^{3/4}$ balls put into n boxes, prob 4 in same box.
4. $\sim n^{(k-1)/k}$ balls put into n boxes, prob k in same box.

Caveat: The approx we used only works when $k \ll n$.
Intent: The above is intended for use when the number of balls is small. What happens when the number of balls is large? Do many boxes get many elements in them?

Recap and Generalize

We state the following informally:

Theorem: Let $n \ll N$. There will be n boxes. There are N balls. The balls are put into the boxes randomly. Then, with high probability, MANY boxes will have MANY balls in them.

Back to Cracking Randomized Shift

September 12, 2019

Cracking Randomized Shift

With a long text Rand Shift is crackable.
If N is long and Eve sees

$$(r_1; \sigma_1)(r_2; \sigma_2) \cdots (r_N; \sigma_N)$$

View as:

1. There are 26 boxes, $\{0, \ldots, 25\}$.
2. Ball i goes into box r_i.

From our study of Bday paradox we know that MANY r's appears MANY times.
Lets see what we can do with one of them: r.

$$(r; \sigma_{i_1}) \cdots (r; \sigma_{i_2}) \cdots \cdots (r; \sigma_{i_L})$$

where L is large.

Cracking Randomized Shift

So we have

$$(r; \sigma_{i_1}) \cdots (r; \sigma_{i_2}) \cdots \cdots (r; \sigma_{i_L})$$

where L is large.

So $\sigma_{i_1}, \ldots, \sigma_{i_L}$ are all coded by the same shift.

Cracking Randomized Shift

So we have

$$(r;\sigma_{i_1})\cdots(r;\sigma_{i_2})\cdots\cdots(r;\sigma_{i_L})$$

where L is large.
So $\sigma_{i_1},\ldots,\sigma_{i_L}$ are all coded by the same shift.

1. From our study of Vig we know that taking every mth letter in a text has the same distribution of letters as a normal text.
2. It turns out that taking a random set of letters also has the same distribution as a normal text.

Good News: Try all shifts and use Is English.
Bad News: Just tells us which shift r maps to.
Good News: MANY boxes had MANY balls so can find many shifts.

Cracking Randomized Shift Final Algorithm

1. Input $(r_1; \sigma_1)(r_2; \sigma_2)\cdots(r_N; \sigma_N)$
2. For each r that appears a lot of time look at where it appeared:
$$(r; \sigma_{i_1})\cdots(r; \sigma_{i_2})\cdots\cdots(r; \sigma_{i_L})$$
3. All of the σ_{i_j}'s used same shift, so find shift like cracking normal shift.
4. We now know what MANY of the r's map to. Should be enough.

Might Help: If know that f is linear then just knowing two r's yields f.

Upshot

1. Det. Ciphers: Message M always maps to the same thing. Boo!
2. We can turn any Det. Cipher into a randomized one. Will use this later in the course.
3. If turn a weak Det. Cipher (like Shift) into a randomized one, still crackable.
4. Cracking it takes a much longer text.