Класични крипто-системи - Криптоанализа

Владимир Божовић

Одсјек за математику
Природно-математички факултет

КРИПТОГРАФИЈА

Још једном о Цезаровом крипто-систему

Као што смо видјели на уводном предавању, можда најстарији крипто-систем је Цезаров алгоритам. Подразумијевамо да свако слово енглеског алфабета има свој бројни еквивалент, из скупа $\{0, 1, \ldots, 25\}$.
Крипто-систем се састоји од функције енкрипције и декрипције, дате као

$$E_k(x) = x + k \pmod{26}, \; D_k(y) = y - k \pmod{26},$$

гдје је k кључ који је такође један од бројева из скупа $\{0, 1, \ldots, 25\}$.

Још једном о Цезаровом крипто-систему

У случају Цезаровог крипто-система, ми заправо вршимо бијективно пресликавање унутар скупа $\{0, 1, \ldots, 25\}$, односно унутар скупа $\{a, b, \ldots, z\}$. На примјер, ако је $k = 6$, онда имамо

$$E_k(x) = x + 6 \pmod{26}, \; D_k(y) = y - 6 \pmod{26},$$

што се рефлектује као бијекција у скупу симбола

$$\begin{array}{ccccc} \mathsf{a} & \mathsf{b} & \mathsf{c} & \cdots & \mathsf{z} \\ \downarrow & \downarrow & \downarrow & \cdots & \downarrow \\ \mathsf{f} & \mathsf{g} & \mathsf{h} & \cdots & \mathsf{e} \end{array}$$

Општи супституциони алгоритам

Цезаров алгоритам је са становишта домена кључа неупотребљив. Пошто кључ условљава једну пермутацију скупа симбола, онда је фактички та пермутација у улози кључа. То нам даје идеју да је заправо могуће узети произвољну пермутацију скупа $\{a, b, \ldots, z\}$ као кључ. На примјер

$$\begin{array}{ccccc} \mathsf{a} & \mathsf{b} & \mathsf{c} & \cdots & \mathsf{z} \\ \downarrow & \downarrow & \downarrow & \cdots & \downarrow \\ \mathsf{s} & \mathsf{k} & \mathsf{h} & \cdots & \mathsf{m} \end{array}$$

је један кључ из скупа у ком има 26! различитих кључева.

Шта ако је домен кључа превелики?

- Видјели смо да је Евин задатак у случају општег супституционог алгоритма знатно тежи, јер не може простом силом (brute force) доћи до правог кључа.
- Треба имати у виду да је $26! = 4 \cdot 10^{26}$ што је огроман број, па се скуп кључева у општем сусптитуционом алгоритму не може проћи ни најсавременијим рачунарим у реалном времену.
- Дакле, Ева мора да користи неку другу технику, односно приступ како би дошла у посјед тражених информација.

Особине природног језика

- Сваки природни језик има неке особине које се могу статистички изразити статистички.
- Учесталост појављивања одређених слова у природном језику, на примјер енглеском, утврђена је одговарајућим таблицама.

Слово	Учесталост
A	0.08167
B	0.01492
C	0.02782
D	0.04253
E	0.12702
⋮	⋮
Z	0.0074

Анализа учесталости

Циљ ове технике је налажење одговарајуће супституционе табеле, односно **кључа** у суспституционом алгоритму. Поступамо по следећој процедури:

- Наћи фреквенције слова у криптограму.
- Упоредити те фреквенције са фреквенцијом слова датом у табели одговарајућег језика основног текста.
- Направити неко смислено повезивање првог слова. Поновити поступак са другим, трећим...
- Пратити "откривање" текста и покушавати комплетирање ријечи. То ће дати нова повезивања у супституционој табели. Ако нема прогреса, вратити се на претходни корак и пробати опет.

Анализа учесталости

- Анализа учесталости даје добре резултате у случају да је криптограм "довољно велики".
- Теоријски, довољно велики криптограм има више од 28 слова, док у пракси, требало би да има најмање 200 слова.

Како се "одбранити" од анализе учесталости?

- Очигледно, један од начина је увећање изворног алфабета, јер упаривање 26 слова није претјерано тежак задатак. Тиме би анализа учесталости би била тежа.
- Омогућити да исто слово основног текста буде енкриптовано различитим словима у криптограму. Тиме се карактеришу полиалфабетски крипто-системи, за разлику од Цезаровог који је био моноалфабетски.
- Претходни задатак можемо остварити тако што омогућимо да енкрипција неког слова зависи од позиције на којој се налази. Остварити такозвану позициону зависност.

Виженеров крипто-систем

Један од најпознатијих полиалфабетских крипто-система је Виженеров крипто-систем. Дуго је третиран као *апсолутно сигуран* система који је коришћен за енкрипцију крајње повјерљивих информација.

- Кључ у Виженеровом систему је ријеч одређене дужине m.
- Испод основног текста се исписује низ кључних ријечи, једна за другом, при чему ће испис последње у низу ићи само до слова које одговара дужини текста за енкрипцију.
- Затим за свако слово појединачно примјенимо Цезаров алгоритам са кључем који одговара слову датим у кључној ријечи испод.

Виженеров крипто-систем

Примјер

Текст за енкрипцију: "SEEYOUINTHEMORNING". Кључ(на) ријеч: "HOPE".
Резултат енкрипције: "ZSTCVIXRAVTQVFCMUU".

S	E	E	Y	O	U	I	N	T	H	E	M	O	R	N	I	N	G
H	O	P	E	H	O	P	E	H	O	P	E	H	O	P	E	H	O
Z	S	T	C	V	I	X	R	A	V	T	Q	V	F	C	M	U	U

Видимо, на примјер, да на крају имамо два слова "UU", што би у Цезаровом алгоритму означавало обавезно енкрипцију истог слова.

Криптоанализа Виженера

- Најприје, јасно је да кључ не смије да буде мале дужине због brute force напада.
- На примјер, ако је кључна ријеч дужине 13, онда имамо 26^{13} кључева, што је отприлике $2.5 \cdot 10^{18}$ кључева.
- Слично, за ријеч дужине 26, имамо $4.2 \cdot 10^{39}$ кључева.
- Дакле, домен кључа није проблем у случају избора довољно дуге кључне ријечи.
- Ипак, мало детаљнији увид у структуру Виженера открива сличне слабости као код Цезаровог алгоритма.

Криптоанализа Виженера

Погледајмо још једном претходни примјер

S	E	E	Y	O	U	I	N	T	H	E	M	O	R	N	I	N	G
H	O	P	E	H	O	P	E	H	O	P	E	H	O	P	E	H	O
Z	S	T	C	V	I	X	R	A	V	T	Q	V	F	C	M	U	U

- Видимо да се слова на позицији 1, 5, 9, 13, ... енкриптују истим словом. Слично је и са позицијама 2, 6, 10, 14, ...
- Ако низ слова које енкриптујемо припада природном језику који има своју статистичку особеност, онда то исто задржава и сваки његов подниз.
- Дакле, слова на наведеним позицијама имају исту дистрибуцију као слова у основном језику.
- Ово нам заправо открива у суштини "слабу тачку" Виженера, односно начин како можемо искористити напад анлизом учесталости.

Криптоанализа Виженера

- Ева ће (највјероватније) бити успјешна уколико дође до **дужине кључне ријечи**.
- Ако је кључна ријеч прекратка, онда отварамо простор за brute force напад.
- Да ли је могуће да дужина кључне ријечи буде (приближно) једнака дужини основног текста?
- Ако је дужина кључне ријечи већа, онда имамо већи број "поднизова" мање дужине, а онда то напад анализом учесталости чини слабијим.

Индекс коинциденције

Дефиниција

Нека је $s = c_1 c_2 \ldots c_n$ низ од n карактера. Индекс коинциденције $IndCo(s)$ је једнак вјероватноћи да два случајно одабрана карактера буду иста.

Теорема

Ако је у низу $s = c_1 c_2 \ldots c_n$ фреквенција слова i једнака F_i, онда је

$$IndCo(s) = \frac{1}{n(n-1)} \sum_{i=0}^{25} F_i(F_i - 1).$$

(Подразумијевамо да се ради о енглеском језику, отуд претходна сума иде од 0 до 25)

Индекс коинциденције

Уколико се ради о енглеском језику, и за довољно велико n, онда се $IndCo(s)$ своди на

$$IndCo(s) \approx \sum_{i=0}^{25} p_i^2,$$

гдје су p_i вјероватноће појављивања појединих слова у енглеском језику. За енглески језик $IndCo(s) = 0.067$.

Теорема

Ако је низ $s = c_1 c_2 \ldots c_n$ настао случајним одабиром из неке азбуке величине m, onda је

$$IndCo(s) = \frac{1}{m}, \text{ што је у енглеском језику } \frac{1}{26} = 0.0385$$

Индекс коинциденције против Виженера

- За претпостављену величину кључне ријечи k, Ева ће узети, на примјер, аритметички низ на позицијама

$$1 + ik,\ i = 0, 1, 2, \ldots.$$

- За тако добијени низ, Ева рачуна његов индекс коинциденције и анализира колико је "удаљен" од језика у којем је написан основни текст.
- Ако се ради о енглеском језику, Евин циљ је да нађе k тако да је индекс коинциденције тог подниза близу 0.067.
- Пробањем различитих вриједности за k, Ева ће послије извјесног броја покушаја врло вјероватно доћи до дужине кључне ријечи.

Касиски тест

Претпоставимо да имамо сајфер-текст за који вјерујемо да је енкриптован помоћу Виженерове енкрипције:

WERXEHJVYSOSPKMUVCOGSIXFUFLTHT**VYC**BTWPTMCLHTRGC
MGQEAGRDVFEGTDJPFPWPGVLIASCSGABHAFDIASEF-
BTVZGIII
HDGID-
DKA**VYC**CXQGJQPKMVIYCLTQIKPMWQEQDYHGEMCTPCK
RAXTKVJSPWVYJXMHNVCFNWRDC-
CMVQNCKXF**VYC**STBIVPDYO
EFBTVZGIIQXWPXAPIHWIC-
SUM**VYC**TGSOPFPLACUCXMSUJCC
MWCCRDUSCSJTMCEYY-
CZS**VYC**RKMRKMVKOJZABXHJFBXGGVR
LIEMKWLTXRDV

Касиски тест

- У нашем примјеру, триграм "VYC" се појављује четири пута, на позицијама: 31, 103, 175, 211 и 253.
- Ово би, на примјер, могла да буде ријеч "THE" у оригналном тексту, а стицајем околности неки фиксни триграм унутар кључне ријеч је ”пао” пет пута испод ње. Зато у криптограму имамо ово понављање "VYC".
- Дакле, дужина кључне ријечи би морала да дијели разлике ”позиција” 103-31, 175-103, 211-175, 253-211, а што је 72, 36, 42.

Касиски тест

- Видимо да су 2, 3 и 6 кандидати за дужину кључне ријечи.
- За сваки од ових кандидата, можемо додатно вршити провјеру помоћу Фридмановог (индекс коинциденције) теста да ли је тестирана дужина кључа заиста исправна, односно да ли је индекс коинциденције поднизова које добијамо близу 0.067.

Шта смо научили из Виженера?

- Велики домен кључа је неопходан, али Виженер је још једна потврда да то није гаранција сигурности.
- Могуће је разбити крипто-систем без претходног проналажења кључа. Дакле, дизајнери морају бити много шири у поимању сигурности од заштите самог кључа.
- Криптограм мора маскирати статистику симбола у основном тексту.
- Ефикасност и сигурност крипто-система су спојени судови, тако да обично појачавање једне карактеристике слаби другу.
- Заштита крипто-система се увијек одвија у референтном систему познатих напада и оклоности, али та се ситуација може брзо промијенити у будућности.

Хвала на пажњи :)