2020 암호분석경진대회 답안제출

2020. 08. 31

참가자 1	**성 명**	이창원
	소속	서울시립대학교 수학과
	휴대폰	
	E-mail	

참가자 2	**성 명**	장호빈
	소속	서울시립대학교 수학과
	휴대폰	
	E-mail	

참가자 3	**성 명**	김인성
	소속	서울시립대학교 수학과
	휴대폰	
	E-mail	

1번 문제 답안

정답:

CRYPTANALYSISISTHESTUDYOFANALYXINGINFORMATIONSYSTEMSINTHESTUDYOFANALYZINGINFORMATIONSYSTEMSINORDERTOSTUDYTHEHIDDENASPECTSOFTHESYSTEMSCRYPTANALYSISISUSEDTOBREACHCRYPTOGRAPHICSECURITYSYSTEMSANDGAINACCESSTOTHECONTENTSOFENCRYPTEDMESSAGESEVENIFTHECRYPTOGRAPHICKEYISUNKNOWNINADDITIONTOMATHEMATICALANALYSISOFCRYPTOGRAPHICALGORITHMSCRYPTANALYSISINCLUDESTHESTUDYOFSIDECHANNELATTACKSTHATDONOTTARGETWEAKNESSESINTHECRYPTOGRAPHICALGORITHMSTHEMSELVESBUTINSTEADEXPLOITWEAKNESSESINTHEIRWEAKIMPLEMENTATIONEVENTHOUGHTHEGOALHASBEENTHESAMETHEMETHODSANDTECHNIQUESOFCRYPTANALYSISHAVECHANGEDDRASTICALLYTHROUGHTHEHISTORYOFCRYPTOGRAPHYADAPTINGTOINCREASINGCRYPTOGRAPHICCOMPLEXITYRANGINGFROMTHEPENANDPAPERMETHODSOFTHEPASTTHROUGHMACHINESLIKETHEBRITISHBOMBESANDBOLOSSUSCOMPUTERSATBLETCHLEYPARKINWORLDWARTWOTOTHEMATHEMATICALLYADVANCEDCOMPUTERIZEDSCHEMESOFTHEPRESENTMETHODSFORBREAKINGMODERNCRYPTOSYSTEMSOFTENINVOLVESOLVINGCAREFULLYCONSTRUCTEDPROBLEMSINPUREMATHEMATICSTHEBESTKNOWNBEINGINTEGERFACTORIZATIONGIVENSOMEENCRYPTEDDATATHEGOALOFTHECRYPTANALYSTISTOGAINASMUCHINFORMATIONASPOSSIBLEABOUTTHEORIGINALUNENCRYPTEDDATATISUSEFULTOCONSIDERTWOASPECTSOFACHIEVINGTHISTHEFIRSTISBREAKINGTHESYSTEMTHATISDISCOVERINGHOWTHEENCIPHERMENTPROCESSWORKSTHESECONDISSOLVINGTHEKEYTHATISUNIQUEFORAPARTICULARENCRYPTEDMESSAGEORGROUPOFMESSAGE

복호화 key: $\begin{pmatrix} 3 & 17 & 12 & 9 & 18 \\ 19 & 24 & 18 & 7 & 12 \\ 11 & 13 & 4 & 12 & 6 \\ 2 & 11 & 9 & 20 & 16 \\ 3 & 21 & 0 & 13 & 23 \end{pmatrix}$ (P=C*K, P:평문, C:암호문, K:복호화 key)

풀이:

Hill cipher는 COA (Ciphertext Only Attacks)에 대해 곧바로 복호화하기 어렵다. 그렇기 때문에 IML (index of maximum likelihood)를 이용하여 전수조사량을 줄인다. Hill cipher에서 key matrix (복호화 key)가 d x d 일 때, matrix multiplication에 대한 fast algorithm을 사용하지 않고 모든 경우에 대한 전수조사 (brute force attack)의 복잡도는 $O(d^3 26^{d^2})$ 이다. Hill cipher는 key의 각 열을 이용하여 계산한 암호문이 다음 암호문에 영향을 주지 않으므로 분할 정복 (divide and conquer attack)을 이용하여 새로운 COA을 적용하면 key의 각 열의 성분이 d개이므로 암호문 각 행 (블록)과 계산 시 $O(d26^d)$, key의 열이 d개이므로 $O(d^2 26^d)$ 의 복잡도를 얻을 수 있다. 그리고 i번째 암호문 행을 C_i, key의 열 x, x'에 대해 x'은 key의 열이 될 수 있는 것 중 사전순으로 x의 앞에 있는 것이라 하자. (예시: $x' = (25,25,...,25,x_t,x_{t+1},...,x_d)$, $x = (0,0,...,0,x_t+1,x_{t+1},...,x_d)$ (25와 0이 각각 t개)) i번째 평문 블록 $p_i = C_i x'$, t를 x의 위에서부터 연속된 0의 개수 $(0 \le t \le d-1)$ 라 하면, $C_i x = p_i + d_{i,t}$ 를 만족하는 $d_{i,t}$에 대해 $d_{i,t} = C_i(x - x')$, x, x'이 사전 순임을 이용하면 $d_{i,t}$은 C_i 의 첫 원소부터 $t+1$번째 원소까지 합의 mod26 값임을 알 수 있다. 이를 통해 $d_{i,t}$값은 우선적으로 구할 수 있으므로, 행렬 곱의 중복된 계산을 줄이면 $O(d26^d)$의 복잡도를 갖는다.

IML은 English alphabet의 frequency를 이용하여 의미가 있을 확률이 높은 plaintext (monogram-wise meaningful string)를 조사한다. IML이 낮을수록 의미가 있을 확률이 높은 것이다. 어떤 문장 P에 대한 IML 값은 $IML(P) = -\sum_i \hat{f}_i \log_2 f_i$, ($\hat{f}_i : P$에서의 $observed\ frequency$, $f_i : (normal)\ letter\ frequency$, $\sum_i \hat{f}_i = 1$) 이다.

1번 문제 답안

주어진 암호문의 총 길이가 1285글자이므로 1285=5*257, 5와 257은 소수이므로 암호문을 5글자씩 257개의 행(블럭)으로 나눈다. 복호화 key 행렬이 5x5 행렬임을 알 수 있고, 복호화 key 행렬의 각 열(x)을 (00000) 부터 $(25\,25\,25\,25\,25)$ 까지 암호문 257개의 행과 계산하며, 위의 $d_{i,t}$ 를 이용하여 중복된 계산을 하지 않았을 때, 각 행을 $O(5\times 26^5)$의 복잡도로 계산할 수 있다. 257개의 암호문 행 (블록)에 대해 계산한 IML이 낮은 5개의 key의 열(x)을 이용하여 120가지의 key에 대해 복호화한다. 그리하여 의미가 있는 평문을 구한다.

위 과정의 의사코드 (pseudo code)는 다음과 같다.

1. $ciphertext\ C$: 길이 $n=md$, 복호화 $key : K$
2. C를 m개의 $block$으로 나눈다. $(C_1,....,C_m)$
3. for $(t=0)$ to $(t=d-1)$:
   for $(i=1)$ to $(i=m)$:
   $d_{i,t}=(C_i$의 첫 원소부터 $(t+1)$번째 원소까지의 합$)(\mathrm{mod}\,26)$
4. K : 임의의 $d\times d$ matrix로 설정
5. I: 길이가 d, 모든 원소를 $-\infty$ 로 설정 // I의 j번째 원소가 K의 j번째 열의 IML 값에 대응된다.
6. for $(i=1)$ to $(i=m)$:
   $p_i=0$ 으로 설정 // p_i는 C_i와 K의 열을 곱한 값
7. $iml=IML(p_1,...,p_m)$ 으로 설정
8. for (사전순의 모든 $d\times 1$의 x $(\mathrm{mod}\,2,\mathrm{mod}\,13$에서 모든 원소가 0인 경우 제외)) :
   $t=(x$의 위로부터 연속된 0의 개수$)$
   for $(i=1)$ to $(i=m)$:
   $iml=iml-\frac{1}{m}log_2 f_{p_i}$ // f_{p_i} : p_i의 frequency
   $p_i=p_i+d_{i,t}\ (\mathrm{mod}\,26)$
   $iml=iml+\frac{1}{m}log_2 f_{p_i}$
   if K의 열 y에 대응되는 IML 값이 iml보다 작을 때 :
   $y\leftarrow x$
   $corresponding\ I\leftarrow iml$
9. K의 열의 순서를 조합하여 복호화한다.

1번 문제 답안

구한 평문은 다음과 같다.

CRYPTANALYSIS IS THE STUDY OF ANALYXING INFORMATION SYSTEMS IN THE STUDY OF ANALYZING INFORMATION SYSTEMS IN ORDER TO STUDY THE HIDDEN ASPECTS OF THE SYSTEMS CRYPTANALYSIS IS USED TO BREACH CRYPTOGRAPHIC SECURITY SYSTEMS AND GAIN ACCESS TO THE CONTENTS OF ENCRYPTED MESSAGES EVEN IF THE CRYPTOGRAPHIC KEY IS UNKNOWN IN ADDITION TO MATHEMATICAL ANALYSIS OF CRYPTOGRAPHIC ALGORITHMS CRYPTANALYSIS INCLUDES THE STUDY OF SIDE CHANNEL ATTACKS THAT DO NOT TARGET WEAKNESSES IN THE CRYPTOGRAPHIC ALGORITHMS THEMSELVES BUT INSTEAD EXPLOIT WEAKNESSES IN THEIR WEAK IMPLEMENTATION EVEN THOUGH THE GOAL HAS BEEN THE SAME THE METHODS AND TECHNIQUES OF CRYPTANALYSIS HAVE CHANGED DRASTICALLY THROUGH THE HISTORY OF CRYPTOGRAPHY ADAPTING TO INCREASING CRYPTOGRAPHIC COMPLEXITY RANGING FROM THE PEN AND PAPER METHODS OF THE PAST THROUGH MACHINES LIKE THE BRITISH BOMBES AND BOLOSSUS COMPUTERS AT BLETCHLEY PARK IN WORLD WAR TWO TO THE MATHEMATICALLY ADVANCED COMPUTERIZED SCHEMES OF THE PRESENT METHODS FOR BREAKING MODERN CRYPTO SYSTEMS OFTEN INVOLVE SOLVING CAREFULLY C ON STRUCTED PROBLEMS IN PURE MATHEMATICS THE BEST KNOWN BEING INTEGER FACTORIZATION GIVEN SOME ENCRYPTED DATA THE GOAL OF THE CRYPTANALYSTIS TO GAIN AS MUCH INFORMATION AS POSSIBLE ABOUT THE ORIGINAL UNENCRYPTED DATAT IS USEFUL TO CONSIDER TWO ASPECTS OF ACHIEVING THIS THE FIRST IS BREAKING THE SYSTEM THAT IS DISCOVERING HOW THE ENCIPHERMENT PROCESS WORKS THE SECOND IS SOLVING THE KEY THAT IS UNIQUE FOR A PARTICULAR ENCRYPTED MESSAGE OR GROUP OF MESSAGE

실행 결과:

```
CRYPTANALYSISISTHESTUDYOFANALYXINGINFORMATIONSYSTEMSINTHESTUDYOFANALYZINGINFORMATIONSYSTEMSINORDERTOST
UDYTHEHIDDENASPECTSOFTHESYSTEMSCRYPTANALYSISISUSEDTOBREACHCRYPTOGRAPHICSECURITYSYSTEMSANDGAINACCESSTOT
HECONTENTSOFENCRYPTEDMESSAGESEVENIFTHECRYPTOGRAPHICKEYISUNKNOWNINADDITIONTOMATHEMATICALANALYSISOFCRYPT
OGRAPHICALGORITHMSCRYPTANALYSISINCLUDESTHESTUDYOFSIDECHANNELATTACKSTHATDONOTTARGETWEAKNESSESINTHECRYPT
OGRAPHICALGORITHMSTHEMSELVESBUTINSTEADEXPLOITWEAKNESSESINTHEIRWEAKIMPLEMENTATIONEVENTHOUGHTHEGOALHASBE
ENTHESAMETHEMETHODSANDTECHNIQUESOFCRYPTANALYSISHAVECHANGEDDRASTICALLYTHROUGHTHEHISTORYOFCRYPTOGRAPHYAD
APTINGTOINCREASINGCRYPTOGRAPHICCOMPLEXITYRANGINGFROMTHEPENANDPAPERMETHODSOFTHEPASTTHROUGHMACHINESLIKET
HEBRITISHBOMBESANDBOLOSSUSCOMPUTERSATBLETCHLEYPARKINWORLDWARTWOTOTHEMATHEMATICALLYADVANCEDCOMPUTERIZED
SCHEMESOFTHEPRESENTMETHODSFORBREAKINGMODERNCRYPTOSYSTEMSOFTENINVOLVESOLVINGCAREFULLYCONSTRUCTEDPROBLEM
SINPUREMATHEMATICSTHEBESTKNOWNBEINGINTEGERFACTORIZATIONGIVENSOMEENCRYPTEDDATATHEGOALOFTHECRYPTANALYSTI
STOGAINASMUCHINFORMATIONASPOSSIBLEABOUTTHEORIGINALUNENCRYPTEDDATATISUSEFULTOCONSIDERTWOASPECTSOFACHIEV
INGTHISTHEFIRSTISBREAKINGTHESYSTEMTHATISDISCOVERINGHOWTHEENCIPHERMENTPROCESSWORKSTHESECONDISSOLVINGTHE
KEYTHATISUNIQUEFORAPARTICULARENCRYPTEDMESSAGEORGROUPOFMESSAGE

[[3, 19, 11, 2, 3], [17, 24, 13, 11, 21], [12, 18, 4, 9, 0], [9, 7, 12, 20, 13], [18, 12, 6, 16, 23]]
```