Cryptic key prediction and statistical inferences thereoff

P. Chakrabarti , *MIACSIT*

***Abstract*—In this paper we have pointed out several ideas based on statistical approaches. The deviation has been analysed and corresponding information can be hence easily computed. We have also shown prediction based on Bayesian belief network model .**

I. INTRODUCTION

Key prediction plays a vital role in case of information transmission from hacker's point of view. As per literature survey[1,2] it is pointed out that key guessing can be done based on summarization principle. Another approach of effective key prediction is use of basic fuzzy operators [3-5].Using comparison analysis of original and predicted values of keys, a hint towards future trend of key values can be achieved and the technique used is neuro-fuzzy model[6,7].Based on data mining principles[8,9] it can also be shown how shared key[10,11] can be generated and extracted in case of multi-party domain.

II. PROCEDURE FOR PAPER SUBMISSION

Let the key space be (k_1, k_2, k_3----k_n), let distribution function $f_1(k_1)$ of random variable k involves a parameter α whose value is unknown and we have to uses value of α on the basis of hacked key space ($k_1,k_2,\ldots\ldots.k_m$) where ($m < n$).

We have to select $\beta = f_2(k_1,k_2,\ldots\ldots..k_m)$, it is basically a number and it is taken as a given for the value of α.

Hence, β is an estimation of α and value of β obtained from hacked key space is on estimate of α.

$|\beta - \alpha|$ should be negligible for successful prediction of key.

Now, we can represent the key hacking criteria as below :

$E(\beta) = \alpha$, for true value of α ---------(1)
and $Var(\beta) <= Var(\Psi)$, for ---------(2)

True value of α and Ψ being any other estimate satisfying equation (1).

Hence the key prediction from hacker's point of view has been pointed out on the basis of property of unbiasedness (equation (1)) and property of maximum variance (equation(2)).

III. PREDICTION BASED ON BINOMIAL DISTRIBUTION

A. Concept

$m_1 = 110111$
$K_1 = 110110$
$K_2 = 000111$
$K_3 = 101011$
$K_4 = 111000$

Game is that sender has one message and it will encrypt the message by any one of the possible four keys using XOR operation. Hacker will know K_1, K_2, K_3, K_4 and encrypted version. Then it will calculate key of highest probability. Then it will perform XOR operation of that key with encrypted message hacked already by him earlier to get back original message. It seems easy as number of keys = 4,but if it is m (m>>1), then computational overhead will be high.

B. . Mathematical Analysis

For a single bit, Bernoulli trial can be applied
where p = probability of correct bit at position i
q = probability of incorrect bit at position i
Since here number of bits = 6, so by binomial distribution we get $p^K q^{n-K}$
where n = number of bits = 6,
p = probability of getting 1 at position 1
q = probability of getting 0 at position 0
K = number of bits of value 1.
The probabilities can be found out based on key value and assigning a frame of 3 bits by its value as 000→ 0.1, 001→ 0.2, and so on.

p of K_1 = min (0.7, 0.7) = 0.7
q of K_1 = 0.3
So, probability of getting $K_1 = 0.7^4*0.3^2$
p of K_2 = min (0.1, 0.8) = 0.1
q of K_2 = 0.9
So, probability of getting $K_2 = 0.1^3*0.9^3$
p of K_3 = min (0.6, 0.4) = 0.4
q of K_3 = 0.6
So, probability of getting $K_3 = 0.4^4*0.6^2$
p of K_4 = min (0.8, 0.1) = 0.1
q of K_4 = 0.9
So, probability of getting $K_4 = 0.1^3*0.9^3$.Find which probability is the largest.

Manuscript received July 31 . 2009.

P.Chakrabarti is with Computer Science and Engineering department , Bengal Institute of Technology and Management , Bolpur-731236 , West Bengal, India (email- prasun9999@rediffmail.com)

So, its corresponding Key will be XORed by hacker with hacked cipher, to get back original message.

IV. PREDICTION BASED ON SERIES ANALYSIS

The most frequent key can be obtained based on Max(f_1,f_2,….,f_n) where f_1, f_2, ….,f_n are relative frequencies and n is total number of keys. *hacked* We can predict the value of a variable if we can measure interval properly. We can apply this scheme in hacking.

Theorem 1

If a variable changes (V) over time (t) in an exponential manner, in that case the value of the variable at the centre point an interval (a_1, a_2) is a geometric mean of its value at a_1 and a_2.

Proof: Let $V_a = mn^a$

Then $V_{a1} = mn^{a1}$ and $V_{a2} = mn^{a2}$

Now, value of V at $(a_1 + a_2)/2$

$$= mn^{(a1+a2)/2}$$
$$= [m^2n^{(a1+a2)}]^{1/2}$$
$$= [(mn^{a1})(mn^{a2})]^{1/2}$$
$$= (V_{a1}V_{a2})^{1/2}$$

In a message there may be a variable which is dependent on any other based on any equation in that case extraction can be made.

Theorem 2

If a variable m related to another variable n in the form m= an, where a is a constant, then harmonic mean of n is related to that of n based on the same equation.

Proof: Let x is no. of given values.

If $m_{HM} = x / (\sum 1/m_i)$ for i = 1 to x

$= x / (\sum 1/an_i)$ [Since $m_i = an_i$]

$= x / (1/a \sum 1/n_i)$ for i = 1 to x

$= a(x / (\sum 1/n_i)$ for i= 1 to x

$= an_{HM}$

V. PREDICTION BASED ON KEY OCCURRENCE CONTROL

Suppose the first hacked key value is k_1 and denote it as minimum boundary (mB), last hacked key value is km and denote it as maximum boundary (MB).

Now, MB and mB are at equal distance from central key k_c ($1 < c < m$), the common difference being 3 times standard duration of k.

Hence, $mB = \ _k - 3 \ \ _k$ -------------(3)

$MB = \ _k + 3 \ \ _k$ ------------(4)

A. Approach 1

Standard value of μ and σ are given μ' and σ' respectively.

Thus, $mB = \mu' - 3* \sigma '/ sqrt(n)$ -------------(5)

[n being total keys in key space]

$mB = \mu' - A \sigma '$ -------------(6)

where $A = 3 / sqrt(n)$

$cB = \mu'$ -------------(7)

Similarly,

$MB = \mu' + A \sigma '$ -------------(8)

where $A = 3 / sqrt(n)$

B. Approach 2

Standard value of μ and σ are not given.

Number of hacking keys = m

For prediction of ith key, Assume that m_i, $\overline{s_i}$ and $\overline{r_i}$ be its respective mean, standard deviation and range

Therefore $\overline{\overline{m}} = \sum_{i=1}^{m} \overline{m_i}$ -------------(9)

$\overline{s} = \sum_{i=1}^{m} \overline{s_i / m}$ -------------(10)

and $\overline{r} = \sum_{i=1}^{m} \overline{r_i} / m$ -------------(11)

Now for estimation process is as follows :

Suppose μ' be $\overline{\overline{m}}$, σ'' be $s / \overline{y}$ (y being function of n)

and r ' be $\overline{r / z}$ (z being function of n)

$mB = \overline{\overline{m}} - 3* \overline{s / (y} * sqrt(n))$ -------------(12)

$cB = \overline{\overline{m}}$ -------------(13)

$mB = \overline{\overline{m}} + 3* \overline{s / (y} * sqrt(n))$ -------------(14)

Similarly, second estimation of σ is as follows :

$mB = \overline{\overline{m}} - 3* \overline{r / (z} * sqrt(n))$ -------------(15)

$cB = \overline{\overline{m}}$

$mB = \overline{\overline{m}} + 3* r / \overline{(z} * sqrt(n))$ -------------(16)

VI. PREDICTION BASED ON PROBABILISTIC APPROACH

Suppose k be the value of hacked keys k_1, k_2, k_3 …………..k_m with respective probability p_1,p_2, …….p_n

When $\sum_{i=1}^{m} p_i = 1$

then $E(k) = \sum_{i=1}^{m} k_i p_i = 1$ -------------(17),

provided it is finite.

Here, we are use bivariate probability based on K ($k_1, k_2, k_3 \ldots\ldots k_m$) i.e. set of hacked keys and
Q ($q_1, q_2, q_3, \ldots\ldots q_n$) i.e. set of predictive keys , ($1 < m < n$)

Theorem 3

If the hacked key set value and predicted key set value be two jointly distributed random variable then

$$E(K + Q) = E(K) + E(Q) .$$

Proof : K assume values $k_1, k_2, k_3 \ldots\ldots\ldots\ldots\ldots\ldots k_m$
Q assume values $q_1, q_2, q_3 \ldots\ldots\ldots\ldots\ldots\ldots q_m$

$$P(K=k_i, Q = q_j) = p_{ij}, \quad i = 1 \text{ to } n \quad \text{and } j = 1 \text{ to } n$$

$$E(K + Q) = \sum_i \sum_j (k_i + q_j)\, p_{ij}$$

$$= \sum_i \sum_j k_i\, p_{ij} + \sum_i \sum_j q_j\, p_{ij}$$

$$= \sum_i k_i \sum_j p_{ij} + \sum_j q_j \sum_i p_{ij}$$

$$E(K + Q) = E(K) + E(Q) \text{ -------------(18)}$$

Similarly, it can be shown that

$$E(K * Q) = E(K) * E(Q) \text{ -------------(19)}$$

VII. PREDICTION BASED ON AUTOREGRESSION PROPERTY

Suppose the hacked key set be K = { $k_1, k_2, k_3 \ldots\ldots\ldots k_m$}. Here from next session onwards, a future key guessing will be done based on

$$k_{m+1} = \varphi + \phi_m k_m + \phi_{m-1} k_{m-1} + \phi_{m-2} k_{m-2} + \ldots\ldots\ldots\ldots + \ldots\phi_1 k_1 + \omega_{m+1} \text{ -------------(20)}$$

Here, ω_{m+1}, indicates a random error at time m+1. Here, each element in the time series can be viewed as a combination of a random error and a linear combination of previous values. Here ϕ_i are the autoregressive parameters.

We can also predict based on the theory of moving average. In the case the key hacking strategy will be based on :

$$k_{m+1} = a_{m+1} + \theta_m a_m + \theta_{m-1} a_{m-1} + \theta_{m-2} a_{m-2} + \ldots\ldots\ldots\ldots\ldots + \theta_{m-q} a_{m-q} \text{ ------------(21)}$$

where a_i is a shock
θ is estimate
q is term indicating last predicted value.

VIII. PREDICTION BASED ON DISPERSION THEORY

The values of the keys for different sessions are not all equal. In some cases the values are close to one another, where in some cases they are highly dedicated from one another. In order to get a proper idea about overall nature of a given set of values, it is necessary to know, besides average, the extent to which the keys differ among themselves or equivalently, how they are scattered about the average.

Let the values $k_1, k_2, k_3 \ldots\ldots k_m$ are the hacked values and c be the average of the original values of $k_{m+1}, k_{m+2}, \ldots\ldots\ldots\ldots\ldots k_n$

Mean Deviation of k about c will be given by

$$MD_c = \frac{1}{(n-m)} \sum_{i=1}^{n-m} | k_i - c | \text{ -------------(22)}$$

In particular , when $c = \overline{k}$, mean deviation about mean will be given by

$$MD_{\overline{k}} = \frac{1}{(n-m)} \sum_{i=1}^{n-m} | k_i - \overline{k_i} | \text{ -------------(23)}$$

IX. PREDICTION USING BAYESIAN NETWORK MODEL

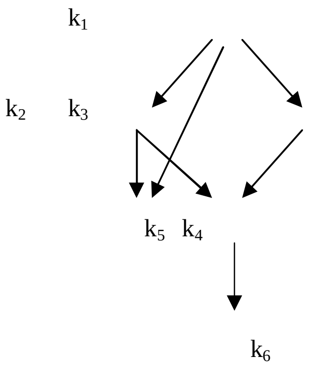

Fig. 1 Bayesian belief network based key prediction

Network ρεπρεσεντατιον for uncertain dependencies are motivated by observations made earlier. We assume that the key space = K = {$k_1, k_2, k_3, k_4, k_5, k_6$} and k_2,k_3 are derived from k_1; k_4 from k_2 and k_3 both ; k_5 from k_1 and k_2 both ; and k_6 from k_4.

Fig. 1 we represent the κεψσ as notes indicating prepositional variables k_i (I=1 to 6), connected by arcs which represents causal influences or dependencies among the keys. The strength of the influences are quantified by conditional probabilities of occurrence of each key. Joint probability of the keys will be given by $P(k_1,k_2,\ldots k_6) = P(k_6 \mid k_4)\, P(k_5 \mid k_2, k_1)\, P(k_4 \mid k_2,k_3) P(k_3 \mid k_1)\, P(k_2 \mid k_1)\, P(k_1)$

IX. CONCLUSION

The techniques involved for key prediction in this paper are namely estimation analysis, binomial distribution, series analysis, key occurrence control, probabilistic approach, autoregression property and dispersion theory. This reveals that key prediction can be done successfully if any relation occurs in case of information transmission thereby reducing the security level. Finally the utility of Bayesian Probabilistic Inference model has been applied for doing the needful prediction.

REFERENCES

[1] P.Chakrabarti et. al. "Approach towards realizing resource mining and secured information transfer" published in international journal of IJCSNS, Korea , Vol 8 No.7, July08

[2] P.Chakrabarti et. al. "An Intelligent Scheme towards information retrieval" accepted for publication in international journal AJIT(Asian Journal of Information Technology), ISSN: 1682-3915, Article ID: 706-AJIT

[3] P.Chakrabarti et. al. " Various New and Modified Approaches for selective encryption with AVK (diffusion and fuzzy) and their comparative study" , published in IEEE Conference , ITNG08, USA , April 08

[4] P.Chakrabarti et. al. "Information representation and processing in the light of neural-fuzzy analysis and distributed computing" accepted for publication in international journal AJIT(Asian Journal of Information Technology), ISSN: 1682-3915, Article ID: 743-AJIT

[5] P.Chakrabarti et. al. "Key generation in the light of mining and fuzzy rule" published in international journal of IJCSNS, Korea , Vol 8 No.9, Sept08

[6] P.Chakrabarti, "Intelligent schemes of neural, curvical cipher generation and congestion control" published in NCSCA-07, ANITS , Vishakapatnam Dec07

[7] P.Chakrabarti ,"Intelligent schemes of cipher generation using comparison analysis and curves", published in International Conference on IT, Jabalpur, Dec07

[8] P.Chakrabarti et. al. "Artificial Intelligence based Cryptosystem" , communicated to an international journal , IJHIS , The Netherlands , August 2008

[9] P.Chakrabarti ,"Analysis of Cryptic data mining", published in International conference on Emerging Technologies and Applications in Engineering, Technology and Sciences , Rajkot ,Jan2008

[10] P.Chakrabarti , "Artificial intelligence based cipher and shared key generation in cryptosystem" selected in ICQMOIT08, Hyderabad , Sep08

[11] P.Chakrabarti et. al. ,"Optimum data transfer and related security", published in National Conference on Methods and models in Computing, Jawaharlal Nehru University,India, Dec07