A Survey of Different Encoding Schemes for Improving the Efficiency of Text based Cryptosystem using ECC

Dhanashree Toradmalle
Assistant Professor
Shah and Anchor Kutchhi Engineering College, Mumbai-400 088.

Saudamini B. Ingale
Student, B.E. IT
Shah and Anchor Kutchhi Engineering College, Mumbai-400 088.

Miheeka G. Chaudhary
Student, B.E. IT
Shah and Anchor Kutchhi Engineering College, Mumbai-400 088.

Aishvarya Akshaya V.
Student, B.E. IT
Shah and Anchor Kutchhi Engineering College, Mumbai-400 088.

Anjali R. Patil
Student, B.E. IT
Shah and Anchor Kutchhi Engineering College, Mumbai-400 088.

ABSTRACT
Providing digital privacy from any kind of malicious activity is referred to as data security. It also includes protection of data from corruption, especially threatened by hackers and eavesdroppers. Various cryptographic systems help provide data security. Elliptic Curve Cryptography(ECC) is advantageous over many cryptographic schemes due to smaller keys and very fast key generation. Smaller key size leads to moderately fast encryption and takes less processing power since computation is less. This paper presents the areas of ECC which have been researched on in depth and a comparative study of various encoding techniques used in ECC implementation.

General Terms
Survey, ECC, encoding schemes.

Keywords
ECC, ECDSA, ECC encryption, encoding schemes, ASCII, matrix mapping.

1. INTRODUCTION
Cryptography is about the design and analysis of mathematical techniques that enable secure communications in the presence of malicious adversaries [1]. A lot of research in the security field has proved public key cryptography to provide excellent confidentiality, data integrity, data origin authentication, entity authentication, and non-repudiation-which are the fundamental objectives of secure communication system[1]. Some public key cryptosystems like RSA and Diffie-Hellman have been widely used for secure data transmission over decades. A richer public key cryptosystem using elliptic curves had been proposed around 1980s. Since then, abundant research done on elliptic curve cryptography has helped in showcasing its efficiency.

The strength of public key cryptosystems relies on the hardness of the mathematical problem used for its construction. RSA is based on integer factorization problem, Diffie-Hellman is based on Discrete Logarithm Problem, whereas the mathematical foundation of ECC is the elliptic curve discrete logarithm problem.

ECC is found to provide equal security with lesser key size than RSA. For example, suppose ECC uses key size of 256 bits, to achieve the same level of data security RSA uses 3072 bits. Elliptic Curve Cryptography (ECC) generates keys with the properties of elliptic curve equation. The elliptic curve equation is:

$y^2 = (x^3 + ax + b) \bmod p$

where a, b, p are the domain parameters of the elliptic curve[2]. Consider the following geometric representation of a general elliptic curve to understand the hardness of mathematical problem on which ECC is based.

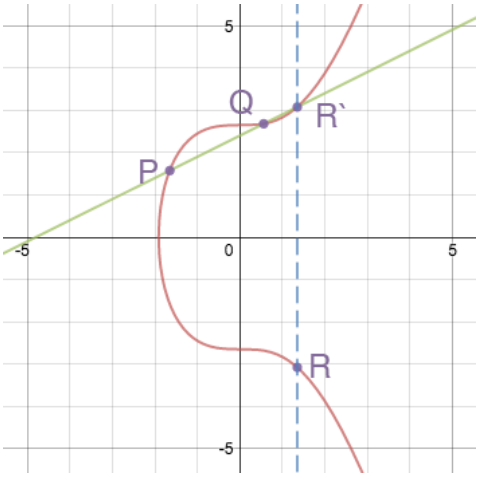

Figure 1. Sample Elliptic Curve

Here P dot Q = R', where P, Q, R, R' are points on the elliptic curve shows what point multiplication means geometrically. The reflection of R' gives R. If the dot operator is used several times or if the point is dotted with itself one ending point is obtained. In ECC, the starting and ending points are kept public whereas the number of times the point is dotted with itself which is 'n' is kept private. Hence it is very difficult to compute 'n' and find the reverse because of which it is called "trapdoor function". Thus, the security obtained using ECC is more compared to RSA or any other public key encryption technique.

The area of digital signatures has been researched on to implement the efficiency of ECC for obtaining strong digital signatures leading to large work on algorithms such as

ECDSA. Also, various attempts done to improve the ECC encryption process of text messages are mainly channelized towards enhancement of the encoding techniques which are used before performing the basic ECC encryption. Phattarin Kitbumrung and Benchaphon Limthanmaphon[3] have unraveled ECC end-to-end encryption to be consisting of four important steps as shown in figure-2. A plaintext message is encoded into a plaintext point followed by encryption. Then the ciphertext is decrypted and decoded from plaintext point into original plaintext message.

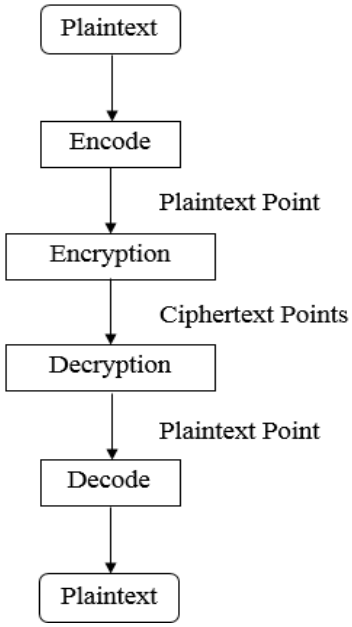

Figure 2. ECC end-to-end encryption[3]

2. LITERATURE SURVEY

2.1 Standards of ECC

Cryptographic operations need to be faster and accurate. To make operations on elliptic curve accurate and more efficient, the curve cryptography is defined over two finite fields:

1. Prime field F(p) and
2. Binary field $F(2^m)$

The equation of the elliptic curve on a prime field F (p) is: $y^2 \bmod p = x^3 + ax + b \bmod p$, where $4a^3 + 27b^2 \bmod p \neq 0$. Here the elements of the finite field are integers between 0 and p – 1. All the operations such as addition, subtraction, division, multiplication involves integers between 0 and p – 1. **Curves with 'p' ranges between 112 to 521 bits.** The domain parameters for the elliptic curve in prime field F(p) are p, a, b, G, n and h.

- **p** is the prime number defined for finite field F(p).
- **a** and **b** are the parameters defining the curve $y^2 \bmod p = x^3 + ax + b \bmod p$.
- **G** is the generator point (**xG, yG**), a point on the elliptic curve chosen for cryptographic operations.
- **n** is the order of the elliptic curve. The scalar for point multiplication is chosen as a number between **0** and **n – 1**.
- **h** is the cofactor where **h = #E(Fp)/n.**
- **#E(Fp)** is the number of points on an elliptic curve.

The equation of the elliptic curve on a binary field $F(2^m)$ is $y^2 + xy = x^3 + ax^2 + b$, where $b \neq 0$. Here the elements of the finite field are integers of length at most m bits. These numbers can be considered as a binary polynomial of degree m – 1. **Curves with 'm' ranges between 113 to 571 bits**. The domain parameters for elliptic curve in binary field are m, f(x), a, b, G, n and h.

- **m** is an integer defined for finite field $F(2^m)$. The elements of the finite field $F(2^m)$ are integers of length at most **m** bits.
- **f(x)** is the irreducible polynomial of degree m used for elliptic curve operations.
- **a** and **b** are the parameters defining the curve $y^2 + xy = x^3 + ax^2 + b$.
- **G** is the generator point **(xG, yG)**, a point on the elliptic curve chosen for cryptographic operations.
- **n** is the order of the elliptic curve. The scalar for point multiplication is chosen as a number between **0** and **n – 1**.
- **h** is the cofactor where **$h = \#E(F2^m)/n$.**
- **$\#E(F2^m)$** is the number of points on an elliptic curve.

The FIPS 186-2 standard recommends elliptic curves over the five prime fields with moduli:[4]

- $p192 = 2^{192} - 2^{64} - 1$
- $p224 = 2^{224} - 2^{96} + 1$
- $p256 = 2^{256} - 2^{224} + 2^{192} + 2^{96} - 1$
- $p384 = 2^{384} - 2^{128} - 2^{96} + 2^{32} - 1$
- $p521 = 2^{521} - 1$.

These primes have the property that they can be written as the sum or difference of a small number of powers of 2. Furthermore, except for p521, the powers appearing in these expressions are all multiples of 32. These properties yield reduction algorithms that are especially fast on machines with word size 32.Curve names: P-192, P-224, P-256, P-384, P-521.Generally curve is defined over F(p) where p has 192 bits, 224 bits, etc. All cofactors (i.e. values of h) for NIST curves are 1 , 2, or 4. All cofactors for prime-field NIST curves are 1

.

Table 1. Standard parameter values of ECC [5]

ECC Parameter Set Name	EA	EB	EC	ED	EE
Bit length of ECC subgroup order *n* (i.e., $\log_2 n$)	160-223	224-255	256-383	384-511	512+
Maximum bit length of ECC cofactor *h*	10	14	16	24	32
Minimum bit length of the hash function output	160	224	256	384	512
Minimum MAC key size (for use in key confirmation)	80	112	128	192	256
Minimum *MacLen* (for use in key confirmation)	80	112	128	192	256

2.2 ECC and Digital Signatures

The research on Elliptic Curve Cryptography has proved its efficiency over other public key cryptosystems. Thus it has been implemented in various applications ranging from simple message transmission between two communicating parties to securing the devices being deployed to achieve 'Internet of Things' (IoT).

The most important application on public key cryptosystem is digital signature. Nowadays, digital signatures are being used to attain integrity, non-repudiation and authentication over the digital data. Understanding the importance of digital signatures in authenticating the users, Abhishek Roy and Sunil Karforma have spelled out the generic phases of digital signature algorithms[6]. They have also explained the basic mechanism of various digital signature schemes such as El-Gamal, DSA, RSA Digital Signature Algorithm, ECDSA, and Elliptic Curve El-Gamal Digital Signature Algorithm by doing an extensive literature survey on digital signatures. These authors have justified the growing acceptance of digital signature schemes using elliptic curves through their paper. A detailed analysis of ECDSA and all its variants along with the security level and execution time of all the phases has been done by Greeshma Sarath, Devesh C Jinwala and Sankita Patel in their paper[7]. Since Elliptic Curve Digital Signature Algorithm (ECDSA) is an elliptic curve variant of digital signature algorithm (DSA), it generates strong digital signatures making use of Elliptic Curve discrete logarithmic problem. They have illustrated the comparison of ECDSA with all its variants in terms of performance and security. Tao LONG and Xiaoxia LIU[8] have worked on enhancing the ECDSA and proposed two improvements to the same. The first improved algorithm suggests removal of mod-inversion operation from the signature generation step of traditional ECDSA, whereas the second improved algorithm suggests the removal of mod-inversion operation from signature generation as well as signature verification steps of traditional ECDSA. This paper throws some light on the comparison of these improved algorithms with traditional ECDSA based on TMU, where TMU is the computing time for two integers multiplied in sense of a mode. The improvements consume shorter time as compared to the traditional ECDSA.

2.3 Various Encoding Schemes

In ECC, only a 'point' on the elliptic curve can be encrypted and decrypted which may be encoded from a message by few schemes such as:

Scheme 'A' - S. Maria Celestin Vigila and K. Muneeswaran [9] have proposed to implement the cryptosystem based on ECC for text based application by transforming the message into an affine point on the elliptic curve over a finite field. In this method, ***each character in a message is first transformed into affine point*** by using the ASCII value of the character as a multiplier to the selected base point of the elliptic curve. After that, the traditional ECC encryption is performed by using the affine point obtained to get the ciphertext; which is then converted back to plaintext character by performing ECC decryption on the receiver side. Refer figure 3.

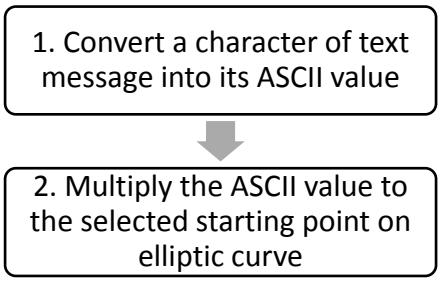

Figure 3. Basic Encoding method

Scheme 'B' - Jayabhaskar Muthukuru and Prof. Bachala Sathyanarayana[10] have put forth mapping techniques for text based messages of fixed and variable lengths using ECC in order to provide multifold security. The paper presents implementation of mapping of a text message into multiple points on elliptic curve with an Initial Vector (IV) using ECC. Here, the text message is first divided into blocks, and each block is EX-ORed with the IV. Then the resultant block is converted into ASCII value of base 256 format. This value is used as a multiplier to the selected starting point of the elliptic curve to transform the message block into an affine point, which is now encrypted into ciphertext using ECC. At the receiver side, the same whole encryption process is applied in reverse manner to obtain the corresponding plaintext message block. In fixed length block mapping technique, the block size is fixed before starting the encryption process, whereas in variable length block mapping technique, each word is considered as a block and null characters are padded to the IV or message block to equate their lengths. The rest of encoding and decoding process is the same in both the techniques. Refer figure 4.

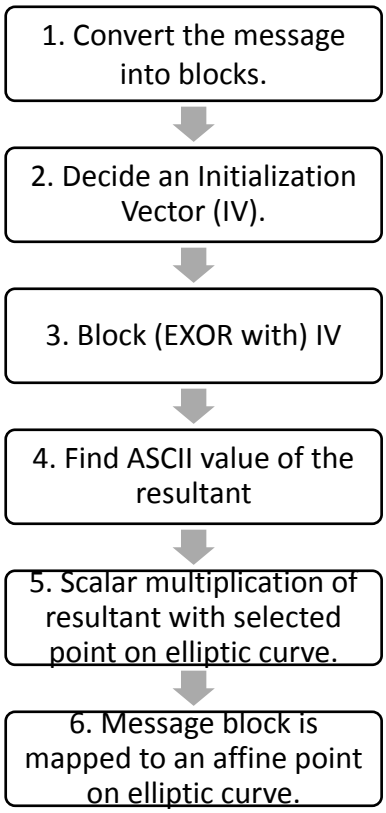

Figure 4. Encoding process in fixed/variable length block mapping technique

Scheme 'C' - Laiphrakpam Dolendro Singh and Khumanthem Manglem Singh[11] have implemented a new technique for text encryption using ECC. In this technique, the text message is converted to ASCII values and these values are divided in groups. Big integers are formed using each group and the group is paired up to be fed as the starting point into ECC encryption. Exactly reverse steps of this process are followed to decode the original text message. Refer figure 5.

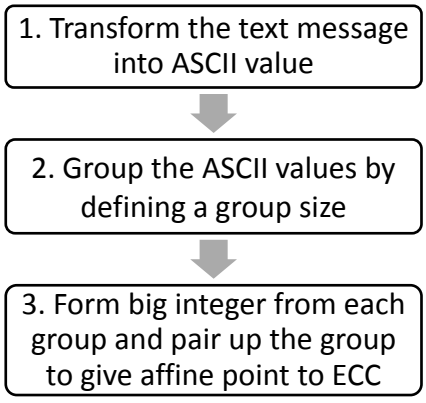

Figure 5. Encoding scheme using grouping method

Scheme 'D' - Balamurugan.R, Kamalakannan.V, Rahul Ganth.D and Tamilselvan.S[12] have studied the matrix based mapping method and combined it with ElGamal technique in ECC. Firstly, the characters in plaintext are mapped to points on Elliptic curve based on non-singular matrix mapping method. Then ElGamal encryption is performed on the mapped points to obtain the ciphertext. On the receiver side, to get back original plaintext, ElGamal decryption is applied and the decoded matrix is multiplied with the inverse of non-singular matrix. The proposal of FGPA implementations of elliptic curve cryptography using matrix mapping concept as well as Steganography using LSB technique in order to develop Crypto-Steg model for security enhancement has been put forward by V.Kamalakannan and S.Tamilselvan[13]. In this method, the text information to be transmitted is first encrypted using ECC based on matrix mapping technique and then it is embedded in cover text to form stego-object.

Scheme 'E' - Phattarin Kitbumrung and Benchaphon Limthanmaphon have surveyed various ECC encoding methods and have classified them into static and dynamic encodings.[3] They have proposed an encoding method based on dynamic point encoding in which the plaintext is mapped onto a point dynamically depending on its ASCII value and sequence ordering.

2.4 Genetic Algorithm & its role

Genetic algorithms are based on evolutionary ideas of natural selection and genetics[14]. The genetic algorithm is an optimization technique that is used to generate more optimal solution. It is inspired by natural evolution mechanism. In a genetic algorithm, a population of strings (called chromosomes), which encode candidate solutions (called individuals or creatures) to an optimization problem, produces better solutions. The genetic algorithm (as shown in figure 6) necessitates a genetic representation of the solution domain and a fitness function to evaluate the solution domain. Usually, these solutions are depicted in binary as strings of 0s and 1s, but other encodings are also possible. The quality of each candidate solution can be evaluated using a fitness function. The evolution usually begins from a population of arbitrarily generated individuals and happens repeatedly in generations. In each generation, the fitness of every individual in the population is evaluated, multiple individuals are indiscriminately selected from the current population based on their fitness, and randomly mutated to form a new population. The new population is then used in the next iteration of the algorithm. Very often, **the algorithm terminates when either an utmost number of generations has been engendered, or a sufficiently good fitness level has been reached for the population**. The individuals in genetic algorithm are typically represented by n-bit binary vectors. The genetic algorithm consists of two basic parameters- one is the crossover probability and another one is mutation probability. Crossover and mutation are two main genetic operators.

Crossover: Different crossover operators help in solving the problem of getting exact copies of parents, if there is no crossover. A specific crossover must be used in order to improve the genetic algorithm problem. Crossover is made so that new chromosomes will contain good parts of old chromosomes and therefore the new ones will be better.

Mutation: Mutation is performed in order to change one or more parts of a chromosome. Mutation predominantly keeps the genetic algorithm from falling into local extremes. Mutation should take place less often, or else GA will become a random search.

The evolutionary algorithms such as genetic algorithm can to be used to find optimal numeric values from search space using heuristic approach.

3. COMPARISON AND ANALYSIS

The encoding techniques explained in schemes 'A', 'B', 'C', 'D' and 'E' are studied from various implementations of ECC. These depict the fact that encoding and decoding steps are the modules where improvements are suggested quite often, to enhance the data security provided by ECC. Few important differences between these schemes are shown in Table 2.

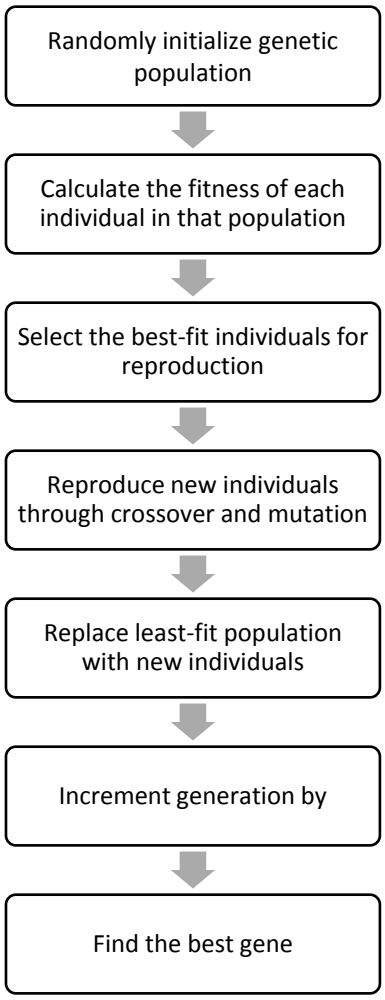

Figure 6. Genetic Algorithm Flowchart

4. CONCLUSION

Because of its advantages over other public key cryptosystems, ECC is found to be an ideal candidate for implementation on constrained devices where speed and memory are limited and low power is available. The encoding schemes help enhance the security level given by traditional Elliptic Curve Cryptography. Although a lot of research is being done in the area of digital signatures and encoding schemes in ECC, still some more research needs to be in depth regarding the cryptosystem based on elliptic curves. The basic differences between few encoding schemes have been discussed in this paper. Depending on the requirement of the application, the suitable encoding method can be utilized for ECC implementation. Also, the genetic algorithms which help find optimal values can be used to optimize various ECC domain parameter. The future scope of this study involves implementation of any of the encoding schemes along with incorporation of basic genetic algorithm to encode a message onto a point on elliptic curve and encrypt it using ECC.

Table 2. Comparison between various encoding techniques

Sr. No	Features	Scheme 'A'	Scheme 'B'	Scheme 'C'	Scheme 'D'	Scheme 'E'
1	**Basic mechanism**	Encoding each character of text message into an affine point on the elliptic curve.	Mapping of each message block onto an affine point using with Initial vector (IV) using ECC.	Grouping the ASCII values of text message which are then paired to map onto a point on elliptic curve.	Each character of text message is encoded into a point on elliptic curve using non-singular matrix.	Every character of plaintext is mapped onto different points over a finite filed dynamically according to its ASCII value and sequence ordering.
2	**Key Factor**	ASCII values are used as a multiplier to obtain the affine	Message block is EX-ORed with IV and then converted into corresponding	Each group is converted into big integer values and paired up to use	Non-singular matrix used in matrix mapping method has only	This scheme is mainly dependent on sequence number of the character in the

		points.	ASCII value to obtain the affine point by using point multiplication.	as starting point for ECC operation.	integer entries.	text message.
3	**Security Level**	Outlines the Brute Force attack on ECC used in wireless networks.	Eliminates exponential time attack due to careful selection of curve.	Largely eliminates Brute force attacks and known plaintext attacks.	Helps prevent Brute force attack.	Ensures protection against frequency cryptanalysis
4	**Advantages**	Single digit ASCII integer of the character is converted into a set of co-ordinates on the elliptic curve thereby removing non-linearity and camouflaging its identity.	Every time the message block maps to different points and thus hides the letter frequencies of the plaintext message.	Avoids the costly operation of mapping and removes the need to share a common lookup table.	Multiple characters are encoded into different points and transmitted simultaneously through matrix mapping.	Every character is mapped to different points depending on ASCII value and its appearing sequence.
5	**Application**	Text based applications	Text based applications	Text based applications	Applications involving messages in form of text and images	Communication in mobile devices

5. REFERENCES

[1] Darrel Hankerson, Alfred Menezes, and Scott Vanstone, "Guide to Elliptic Curve Cryptography", Springer: 2014, pp. 2-3.

[2] Samta Gajbhiye, Monisha Sharma, and Samir Dashputre, "A Survey Report on Elliptic Curve Cryptography", International Journal of Electrical and Computer Engineering (IJECE): 2011, pp. 195.

[3] Phattarin Kitbumrung and Benchaphon Limthanmaphon, "ECC Dynamic Point Encoding on Mobile Device", IEEE: 2011.

[4] "Recommended Elliptic Curves for Federal Government", pp. 24-28.

[5] Elaine Baker, Don Johnson and Mikes Smid, "NIST Computer Security: Recommendation For Pair-Wise Key Establishment Schemes Using Discrete Logarithm Cryptography", March 2011, pp. 29.

[6] Abhishek Roy and Sunil Karforma, "A survey on digital signatures and its applications", J. of Comp. and I.T. Vol. 3(1&2): 2012, pp. 45-69.

[7] Greeshma Sarath, Devesh C Jinwala and Sankita Patel, "A Survey on Elliptic Curve Digital Signature Algorithm and Its Variants", CS & IT-CSCP: 2014, pp. 121-136.

[8] Tao LONG and Xiaoxia LIU, "Two Improvements to Digital Signature Scheme Based on the Elliptic Curve Cryptosystem", International Workshop on Information Security and Application (IWISA): 2009.

[9] S. Maria Celestin Vigila and K. Muneeswaran., "Implementation of Text based Cryptosystem using Elliptic Curve Cryptography", ICAC: 2009.

[10] Jayabhaskar Muthukuru and Bachala Sathyanarayana, "Fixed and Variable Size Text Based Message Mapping Techniques Using ECC", Global Journal of Computer Science and Technology, Volume 12, Issue 3, Version 1.0: February 2012.

[11] Laiphrakpam Dolendro Singh and Khumanthem Manglem Singh, "Implementation of Text Encryption using Elliptic Curve Cryptography", IMCIP: 2015.

[12] Balamurugan.R, Kamalakannan.V, Rahul Ganth.D and Tamilselvan.S, "Enhancing Security in Text Messages Using Matrix based Mapping and ElGamal Method in Elliptic Curve Cryptography", (IC3I) IEEE: 2014.

[13] V. Kamalakannan and S. Tamilselvan, "An Efficient Crystography Protocol Using Matrix Mapping Technique", IEEE ICCSP: 2015.

[14] Richa Garg and Saurabh Mittal, "Optimization by Genetic Algorithm", International Journal of Advanced Research in Computer Science and Software Engineering: April 2014, pp. 587-589.