Journal of Advances in Science and Technology
Vol. 21, Issue No. 2, September-2024, ISSN 2230-9659

DOI: https://doi.org/10.29070/x7qnew50

Check for updates

A Study the DNSSEC Cryptography Technique to implement data security

Mausami Arya [1] * , Dr. Divyarth Rai [2]

1. Research Scholar, LNCT, Bhopal, Madhya Pradesh, India
mausamarya16@gmail.com ,
2. Professor, Department of Computer Science, LNCT, Bhopal, Madhya Pradesh, India

Abstract: The research was conducted using patient records or data received from a cloud medical resource. For DNS security to work, medical decision-makers must be able to access server data. Through DNSSEC in data security, DNS security can be obtained in existing DNS. Implementing DNSSEC switches from standard security algorithms improves security, but it leads to challenges when dealing with IP fragmentation & prone to DDOS attacks. If DNSSEC isn't optimised for request distribution using vulnerable to IP fragmentation and powerful DDOS attacks using ECC. Although there are some drawbacks, elliptic curve cryptography (ECC) alternative cryptosystems to RSA are intriguing and could be useful for DNSSEC. Future research should look into the feasibility of implementing ECC on a large scale. The DNSSEC Cryptography algorithm will ensure the DNS heterogeneous medical data server is authentic & secure.

Keywords: DNS Security, Medical Data, Cryptography, Elliptic Curve Cryptography, RSA

INTRODUCTION

The DNSSEC cryptography technique should be used to implement data security. Implementing ECC with RSA for data security is possible in the existing DNS Security. This study proposes a novel architecture for implementing cryptographic algorithms using DNSSEC, making systems more resistant to DoS and IP fragmentation. In our data mining analytical process, we have integrated cryptographic features.

After encrypting data saved in the cloud, it can be used for similarity matching & data retrieval. The Elliptic Curve Diffie-Hellman algorithm can decode heterogeneous data stored in different DNS systems. What follows is the crucial question in this case. For what reasons is it important to keep the medical data server secure? The disparate datasets gathered from various domains' medical histories. To access server data via medical decision makers, DNS security is necessary. Through DNSSEC in data security, DNS security can be obtained in existing DNS. While DDOS attacks are possible or IP fragmentation presents difficulties, using DNSSEC changes from traditional security methods increases security.

According to Roland Van R.D. (2017), DNSSEC is responsible for enhancing DNS security. A number of issues have been brought to light regarding DNSSEC. The addition of digital signatures to the DNS by DNSSEC causes DNS answers to grow in size. since of this, DNSSEC is a prime target for amplification-based denial-of-service attacks since it is more vulnerable to packet fragmentation. Crucial management policies can also be somewhat intricate. Due to this, DNSSEC becomes vulnerable and can have operational issues.

Elliptic curve cryptography allows for powerful distributed DoS attacks against DNSSEC, which is

vulnerable to IP fragmentation; artificial bee colony optimisation is useful for optimising request distribution, which improves efficiency. Although elliptic curve cryptography-based alternatives to RSA do exist, they see very little application in DNSSEC. These offer both advantages and problems, but they are ideal for DNSSEC. Studies that attempt to determine if large-scale ECC deployment in DNSSEC is feasible. Verify the DNS heterogeneous medical data server's legitimacy and integrity using the DNSSEC cryptography technique.

BACKGROUND KNOWLEDGE

Elliptic Curve Diffie Helman

(P. Hoffman 2012) Digital signatures, key encryption & decryption, and the PKC's foundational principles the public and private keys enable it to solve the discrete logarithm problem (DLP). Many PKC systems currently employ the RSA algorithm system. An elliptic curve is a fundamental mathematical topic that belongs to the domains of geometry and algebra. To alleviate the computational burden on the DLP, the PKC made use of elliptic curves, which were presented by Miller and Koblitz in 1985. These curves encompass key exchange, encryption, and digital signatures.

Elliptic Curve Cryptography

ECC is a way to encrypt public keys. Elliptic curves are essentially two-variable cubic equations with coefficients. The variables, locations, and coefficients used by ECC are all elliptic curves, and they are all limited to elements of a finite field.

In 2017, Roland Van R.D. Because ECC uses elliptic curve cryptography as its foundation, the variables and coefficients it employs can only be elements of finite fields, where each ingredient is contained in a limited quantity. Given the frequency with which operations such as addition & multiplication are performed, the finite field can include any two elements. The fields are ordered by their size, and each one is unique up to isomorphism; the size of this field is pn, where p is a prime number and n is a positive integer. The mathematical form of a series represented as p in a finite field is Fp. A couple of varieties of ECC:

- A basic curve in the finite field Fp, the elliptic curve can be described as 0, 1,..., p-1.
- The elliptic curve is a binary curve in the finite field F2m, where m is a big integer and F2m can be expressed as 0, 1,..., 2n-1.

The current study relies on the prime field Fp since ECC prime curves outperform classical cryptography in terms of efficiency and security (G. van den Broek 2014). The standard method to represent an elliptic curve E is as follows: y2 = x 3 + ax+b. If we take an elliptic curve E and plug in any two points (x, y) in the prime field Fp, we get an equation that applies.

$$y^2 = x^3 + ax + b \pmod{p} \quad (1)$$

To formulate the previous equation as Ep(a, b), we use the fact that x, y, a, and b are elements of the finite field Fp when p is a big prime number. A positive value for b is also necessary for this equation to hold.

$$4a^3 + 27b^2 \,(\text{mod}p) \neq 0 \quad (2)$$

Additionally, the elliptic curve includes an O-shaped point at infinity. An elliptic curve's order n is equal to the total number of points, including infinity, on the curve. To satisfy the condition, Ep(a,b) must be equal to zero.

$$p+1-2\sqrt{p} \leq n \leq p+1+2\sqrt{p} \quad (3)$$

In ECDH, the laws of addition are:

- The additive's identification serves its purpose. P on the elliptic curve can be expressed as P+0=P.
- If P =(X.Y) and –P =(X.-Y). then P+(-P) = P-P=0
- On the elliptic curve (a,b), where P is the point between xp & yp and Q is the set of xQ & yQ, the point R is the place where the straight line PQ intersects the elliptic curve (a,b) if and only if P+Q=-R.
- Two points P added together form P+P=2P. The tangent line through point P is 2P.

Here are a few examples of using multiplication tables:

Adding P to itself k times is the definition of multiplying a point P on an elliptic curve by an integer k.

$$Pxk = (P+P+ \ldots..+P) \quad (4)$$

The ECC key-exchange system is reliable. Elliptic Curve Diffie Hellman (ECDH) ECC is a key exchange mechanism that generates public and private keys for a shared key. From the collection of public parameters Ep(a,b), ECDH selects a base point G = (x, y) starting with a large prime number P. If there exists a large integer n with a base point order such that nG=0, then G and Ep(a,b) are both prime numbers. Here we will look at a significant communication between two users as an example to demonstrate the procedure.

1) In order to create a public key PA= nA×G, which is associated with Ep(a,b), User A chooses a private key that is less than the order n.
2) Ep(a,b) is associated with the public key PB= nB×G, which is generated after User B chooses a number less than n for its private key.
3) Once the public keys have been exchanged, user A will generate their private key $KA = nA \times PB$ and user B will generate their private key . In $KB = nB \times PA.$ this case, KA is the same as KB since

$$K_A = n_A x P_B = n_A x (n_B x G) = n_B x (n_A x G) = n_B x P_A = K_B \quad (5)$$

Considering that it can provide the same degree of security as the RSA method while using a smaller key complexity, ECDH has been utilised in numerous studies investigating wireless sensor networks to create

information security solutions. As a result, wireless sensor networks are ideal environments for ECDH.

Whether ECC is secure is decided by the results of DLP. It is sufficient to compute an additional point Q = kG using the given k & G, assuming G is the base point on Ep(a,b) of order n, where Q ϵEp (a,b) and k < p. Calculating k in a limited amount of time when Q and G are known is extremely challenging because n is so huge. A 1024-bit RSA key is a good starting point for DLP security, and the To prevent cracking, the elliptic curve equation needs a prime number P that is bigger than 160 bits.

A prime number greater than or equal to 160 bits (p≠n) is required for the elliptic curve's base points n. Ronald Van R.D. (2017) states that in order to attain the same degree of security as 1024-bit RSA, an ECC key length of 160 bits is necessary. The reduced calculation complexity & storage requirement of ECC make it an attractive choice for key management. ECC keys are 160 bits long & use ECDH for key exchange.

Support Vector Machines

Over the past decade, SVMs have seen widespread use in data mining or ML, and they have subsequently discovered new applications in a wide variety of fields. Support vector machines (SVRs), ranking SVMs (or RankSVMs), & categorizing SVMs are the terms used to describe SVMs when they are trained to learn regression, ranking, or classification tasks, respectively. A particular SVM achieves two unique qualities.

1. High degree of generalisability through margin maximisation.
2. The second point is to back the use of the kernel method for effective learning of nonlinear functions.

Classification, regression, and ranking function learning with SVMs are introduced in this chapter, along with these broad principles and methods. Specifically, SVMs for binary classification are introduced initially.

SVM CLASSIFICATION

With their origins in classification, support vector machines (SVMs) have since been expanded to include regression and preference (or rank) learning. When first implemented, SVMs take the form of a binary classifier, where the learnt function's output is a positive or negative value. By utilising the pair-wise coupling method to combine diverse binary classifiers, multiclass classification can be achieved (P. Hoffman 2012, Javier Rodriguez 2017). Here, we break down SVM's background and formalisation as a binary classifier, in addition to its two most essential characteristics, which are the kernel trick & maximisation of the margin.

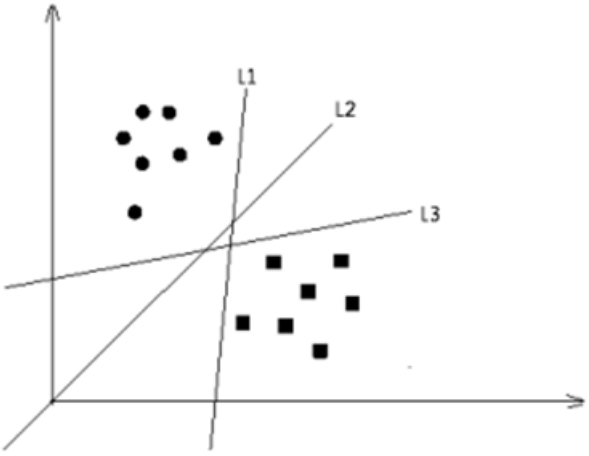

Figure 1: Two-dimensional hyperplane linear classifiers

The ability to distinguish between two groups is possessed by binary SVM classifiers. One n-dimensional vector represents each details object (or details point). There are only two possible categories to which each of these details belongs. They are separated by a hyper plane in a linear classifier. Take Fig. 1 as an example; it depicts two sets of details in two-dimensional space, separated by hyper planes and specific boundary lines. In Figure 1, we can see that there are a number of linear classifiers that can accurately categorise the data into the two sets labelled L1, L2, and L3.

When SVM uses the hyper plane with the largest margin, it is able to achieve the maximum separate the two classes. A pair's margin is the total of the shortest paths, as measured from the hyperplane that separates them, to the bidets. In terms of generalisability, hyperplane specificity, and the ability to accurately categorise "unseen" or test-specific features, the former two are more likely to be true. With the use of SVMs, nonlinear classification problems can be supported by doing the inverse mapping of input space to feature space.

The kernel approach is useful for this purpose since it allows for a lack of accurate specification of the mapping function, which could lead to the curse of dimensionality issue. Given that both the input and feature spaces have been changed, a linear classification in the former is equivalent to a nonlinear classification in the latter. SVMs achieve this by constructing a hyper plane with the maximum separation from the input vectors & positioning it in a higher-dimensional space, which is also known as feature space.

Hard-Margin SVM Classification

As an example of a hard-margin SVM configuration, which happens when the training data is blast-free and can be adequately categorised by a linear function, it helps clarify how SVMs find the hyper plane of maximum margin or enables nonlinear classification. Using mathematics, we can express points D, which stand for training decisions, in the following way:

$$D = \{(x_1, y_1),(x_2, y_2),\ldots,(x_m, y_m)\} \quad (6)$$

A real vector with n dimensions, xi, and a value of 1 or -1 for yi that indicates the class to the highest

possible value of xi. Classification via support vector machines (SVM) F(x) looks like this.

$$F(x) = w \cdot x - b \quad (7)$$

Simple V-shaped weights (W) and biases (b) are calculated by SVM during teaching. F(·) (or w + b) should return positive values for specific points that are positive and negative numbers else for each point xi in D., w· xi −b > 0 if yi = 1, and w· xi −b < 0 if yi = −1, for accurate training data classification. It is possible to transform these terms into

$$y_i(w \cdot x_i - b) > 0, \ \forall (x_i, y_i) \in D \quad (8)$$

If a single linear function F correctly classifies all points in D or if F satisfies Equation, then D is said to be linearly separable. Additionally, the margin should be maximised by F, also known as the hyperplane. The distance from the hyperplane to the nearest points of detail is determined by the margin. Let us example of a unique hyperplane. This is accomplished by modifying Eq. into the following Eq.

$$y_i (w \cdot x_i - b) \geq 1, \ \forall (x_i, y_i) \in D \quad (9)$$

Note that the right side of one equation becomes 1 instead of 0, and there's an equality sign in that equation. A single F satisfying Eq exists if and only if D is linearly separable or if each point in D satisfies Eq. This is because, provided that w and b exist, rescaling them to satisfy Eq is always an option. A vector xi's inverse distance from the hyper plane is given by the formula |F(xi)| ||w||.

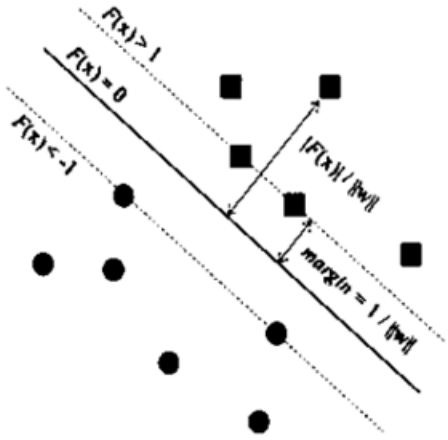

Figure 2: The support vector machine classification function: the two-dimensional hyperplane that maximises the margin

Soft-Margin SVM Classification

We have only covered linearly separable instances up to this point in the discussion. Nevertheless, the optimisation problem and all of its expansions will remain unsolved unless D is linearly separable. Soft margin SVM maximises the margin while allowing mislabeled details points, which is useful for some scenarios. As an indicator of the level of misclassification, the technique introduces slack variables, ξi. A soft margin SVM optimisation problem is on the horizon.

Be sure that the dual problem does not contain the slack variables ξi or their Lagrange multipliers. With

one small but significant exception, the dual problem for the absence of separable designs is the same as the simple case for linearly separable designs. The function that requires to be maximised, Q(α), remains identical in both cases. The situation when there is no separable data is similar to the separable scenario, but with a more strict requirement of C ≥ αi ≥ 0, instead of the less strict αi ≥ 0. After this adjustment, the biassed b and weight vector w optimisation, that is restricted to the linearly separable situation, remains unchanged.

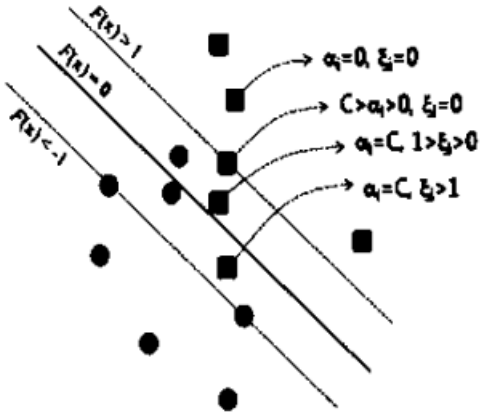

Figure 3: Graphic relations among αi , ξi , along with C

Kernel Trick for Nonlinear Classification

It is impossible to distinguish between classes on a straight hyperplane should the data used for training not be linearly separable. To train a nonlinear function in some instances, it is required to expand linear SVM to nonlinear SVM for the classification of details that cannot be separated in a linear fashion. Finding classification functions with nonlinear SVM is a two-step approach. A prerequisite to linearly separating the training data is the transformation of the input vectors into high-dimensional feature vectors.

The following stage for support vector machines using the newly generated feature space is to locate the hyper plane with the highest margin. In the updated feature space for the separation hyper plane, a function that was nonlinear in the initial input space becomes linear. Keep in mind that the kernel function maximises its output when the two vectors are equal; this is because it is a type of identically function among two vectors. This means that SVMs can learn to compute an inverse function for any pair of details objects, regardless of the shape of the details (such as a tree or graph). We will not go into further detail about the characteristics of these kernel functions here.

IMPLEMENTATION

Data security in the context of clinical decision support systems requires DNS security measures in cloud resources.

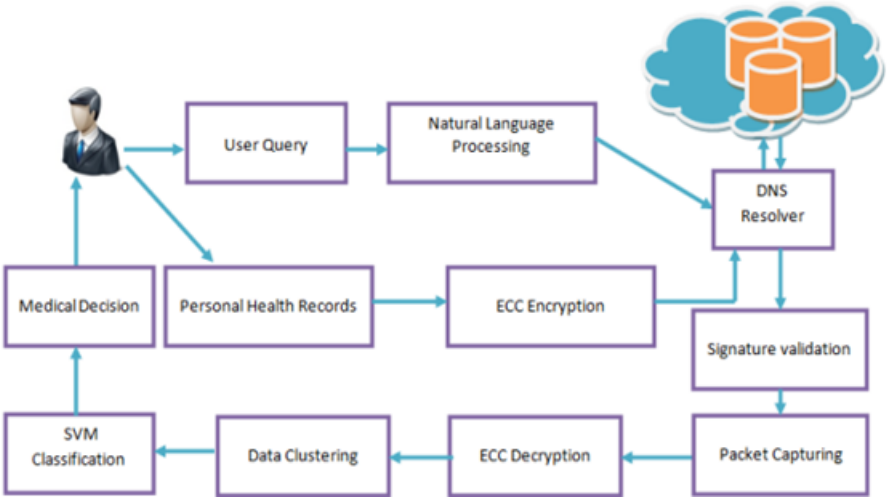

Figure 4: Proposed Architecture

Patients who have questions about their medical decisions can use this system through the CDSS. This means that the query results can be in the form of a sentence or patient-specific parameters like medical history. In order to identify and remove stop words, Natural Language Processing is employed to extract stem words from the ongoing text processing process that begins with splitting. After identifying potential ambiguities, the following step is to remove them using word meaning disambiguation. In order to validate the user's signature with their details, DNS Resolver receives the discovered stem words and uses them as search terms. While in training, you can also add the patient health records that were collected from the Clinical Decision Support System to the DNS servers so they can be used later on. Whoever else is making decisions could find this material helpful. Data encryption and signature resolution are both handled by ECC in this case. We must ensure that the stored data can only be accessed by approved individuals. Data can only be accessed by authorised users using the public key and private key. The DNS Resolver can validate the user's signature by comparing it to the request's signature and then capturing the packet. The data packets that were retrieved should be decrypted using the ECC Decryption methods.

DNSSEC CRYPTOGRAPHY

An IP address-to-domain name mapping technique is known as the Domain Name System (DNS). To translate between common domain names (such as www.uptu.ac.in) and numerical IP addresses (such as 117.211.115.134) in IPv4 or IPv6 networks, the DNS uses a hierarchical, distributed database. Every node in the DNS tree stands for a DNS name. At the node's level, there is a DNS domain. By converting hostnames to their numerical equivalents, the network can follow a different path to a certain server than a customer would. Networks unique to the internet rely on accurate and timely DNS translations, making them an attractive target for attackers. Originally, DNS did not have a way to indicate if a domain name's data were faked or belonged to an authorised owner. Due to this security hole, the system is susceptible to several attacks, such as DNS spoofing and cache poisoning. An attacker can inject a malicious record into the DNS details base if they are able to predict a DNS message ID and reply before the legitimate DNS server. This is possible due to weak authentication among DNS servers sharing updates.

In order to map the erroneous host to an IP address, the exploit has a hacked DNS server request the

attacker's DNS server. A decision of the Internet Engineering Task Force (IETF) staling withers, DNS Security Extensions (DNSSEC) aim to fix DNS vulnerabilities and safeguard against cyber threats. In order to strengthen Internet security in general, DNSSEC primarily aims to address and resolve DNS security vulnerabilities. An essential component of DNSSEC is the authentication feature it provides to DNS, which significantly increases system security.

The three subsequent IETF Requests for Comments that were released in March 2005 laid out the DNSSEC fundamental parts are RFC 4033 is DNSSEC Overview along with Necessities, RFC 4034 Resource Records for the DNSSEC and RFC 4035 Protocol Modifications for the DNSSEC Existing.

Approaches to DNS security primarily rely on public-key cryptography. The DNSSEC authentication public key architecture is based on the Rivest Shamir Adleman Algorithm (MD5/RSA) & Digital Signature Algorithm (DSA). The main benefit of digital signatures made using the public key architecture is that they may be verified by anyone in possession of the public key. The basic principle is that all nodes in the DNS space should have public keys and that all messages sent to DNS servers should be encrypted with private keys.

What is the process for creating certificates and signatures to merge the identity data of top-level domains, given that DNS is public and authenticated DNS roots public keys are known to everyone? All the children of a patient in DNS have their public keys signed by that patient.

The initial design of DNS did not include any way to indicate if domain name information were forgeries or belonged to an authorised domain owner. For example, DNS spoofing and DNS cache poisoning are two types of attacks that can exploit this security hole. An attacker can inject a malicious record into the DNS details base if they are able to predict a DNS message ID and reply before the legitimate DNS server. This is possible due to weak authentication among DNS servers sharing updates.

An attacker might potentially inject cache servers with bogus information, which is one of the new DNS server architecture vulnerabilities. In order to map the erroneous host to an IP address, the exploit has a hacked DNS server request the attacker's DNS server. A DNS cache poisoning attack occurs when an attacker inserts a malicious IP address into a DNS database, causing it to be cached instead of genuine. Along with malicious software (such as a worm, spy ware, browser hijacker, etc.) and requests for legitimate addresses being redirected, the buyer's machine can actually download these threats from the malicious website. To guarantee that lookup details are accurate and that connections adhere to legitimate servers, DNSSEC uses cryptographic keys and digital signatures. Pictured in figure 5 is a DNS spoofing scenario: in order to get an IP address, the client includes its ID in the query it sends to the DNS server. If an attacker is able to identify a client's ID, they can send harmful details or redirect them to a malicious site.

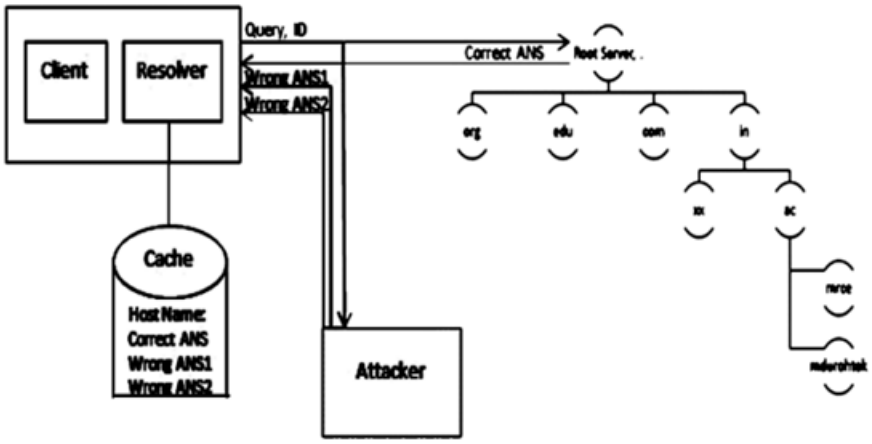

Figure 5: DNS Spoofing

The two pillars of DNS security are details authentication (the information's authenticity) and details integrity (the fact that it has not been altered). Since DNS is publicly available, there is no need to worry about confidentiality. To strengthen up the DNS and make it more resistant to cyberattacks, the Internet Engineering Task Force (IETF) settled on DNS Security Extensions (DNSSEC) as a staling valve. By fixing DNS security flaws, DNSSEC aims to make the Internet a safer place for everyone.

TECHNIQUES

Symmetric Key Cryptography

The pair encryption & decryption processes share a common habited key in this cryptography type. Before beginning data exchange, the key must be securely habituated between the two sides. This is both faster and easier to implement. The following diagram does a good job of illustrating the idea.

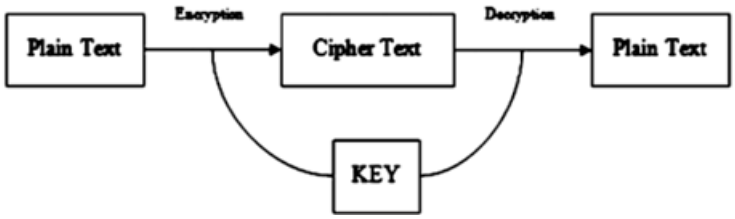

Figure 6: Symmetric Key Cryptography

Public Key Cryptography

Shabbier details necessitate a critical pair. The encryption public key and the decryption private key are both known to all; the buyer alone possesses the private key. Because it takes a lot of effort to make the key pair, its design is sluggish. Here is a diagram that should help you understand the concept.

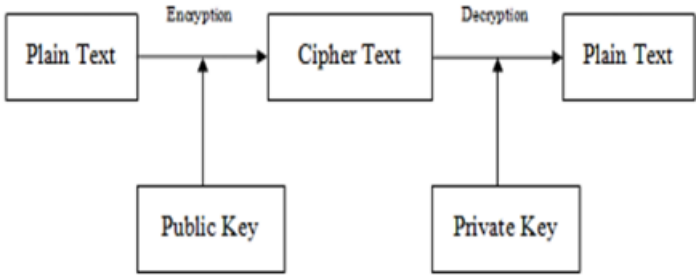

Figure 7: Public Key Cryptography

Elliptic Curve Cryptography

Elliptic curve cryptography (ECC) is a way to encrypt public keys.

CONCLUSION

DNSSEC in action is shown here retrieving patient records pertaining to medical cardiac arrests from several DNS data servers. Implementation and utilisation of DNS resolvers for data cryptography techniques have been successful. Classifying medical judgements regarding cardiac arrest disease is done using the obtained encrypted data. In the future, Real-time vulnerabilities involving IP fragment & DoS attacks have validated DNS Security and its applying the concepts of authenticity & integrity.

References

1. Abdeldjouad, F, Brahami, M & Matta, N 2020, 'A Hybrid Approach for Heart Disease Diagnosis and Prediction Using Machine Learning Techniques', The Impact of Digital Technologies on Public Health in Developed and Developing Countries, Lecture Notes in Computer Science, vol. 12157, pp. 299

2. Barakat, H, Andrew, P, Bradley & Mohammed Nabil Barakat, H 2009, 'Intelligible Support Vector Machines for Diagnosis of Diabetes Mellitus', IEEE Transactions on Information Technology in Bio Medicine Journal, vol.14, no. 4, pp. 1-7

3. Chitra, R., & Seenivasagam, V. (2013). Review of heart disease prediction system using data mining and hybrid intelligent techniques. ICTACT journal on soft computing, 3(04), 605-09.

4. David Wai-Lok Cheung, Vincent T. Ng, Ada Wai-Chee Fu, and Yongjian Fu, Efficient Mining of Association Rules in Distributed Databases, IEEE Transactions on Knowledge and Data Engineering, Vol. 8, No. 6, pp. 911-922, December 1996

5. Folorunsho O. (2013) Comparative Study of Different Data Mining Techniques Performance in knowledge Discovery from Medical Database , International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 3, Page no.11

6. Ishaq, A., Sadiq, S., Umer, M., Ullah, S., Mirjalili, S., Rupapara, V., & Nappi, M. (2021). Improving the prediction of heart failure patients' survival using SMOTE and effective data mining techniques. IEEE access, 9, 39707-39716.

7. Krishnaiah, V., Narsimha, G., & Chandra, N. S. (2016). Heart disease prediction system using data mining techniques and intelligent fuzzy approach: a review. International Journal of Computer Applications, 136(2), 43-51.

8. Patil, S. B., & Kumaraswamy, Y. S. (2009). Extraction of significant patterns from heart disease warehouses for heart attack prediction. IJCSNS, 9(2), 228-235.

9. Roland Van R.D, Kaspar Hageman, Anna Sperotto and Aiko Pras, "The Performance Impact of Elliptic Curvve Curve Cryptography on DNSSEC Validation", IEEE/ACM Transactions on Networking, vol.25, No.2, April 2017, pp.738-750.

10. Soni, J, Ansari, U, Sharma, D & Soni, S 2011, 'Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction', International Journal of Computer Applications vol.17,no.8, pp. 43-48

11. Su, YS, Ding, TJ & Chen, MY 2021, 'Deep Learning Method in Internet of Medical Things for Valvular Heart Diseases Screening Systems, IEEE 101 J., vol. 8, pp. 16921-16932.

12. Tougui, I., Jilbab, A., & El Mhamdi, J. (2020). Heart disease classification using data mining tools and machine learning techniques. Health and Technology, 10(5), 1137-1144.