Preserving Confidentiality and Privacy of Sensitive Data in e-Procurement System

Rajesh Narang¹ Chief Technology Officer, National e Governance Division, New Delhi, India
Tanmay Narang², Ex-IT-Analyst, KPMG, DLF Phase 3 Cyber City, Gurgaon, India.
¹rajesh.narang@digitalindia.gov.in, ²tanmay.narang92@gmail.com

ABSTRACT

With the inclusion of direct purchases by Government in e-Procurement System, several security agencies are joining it but they expect their sensitive data to remain hidden. So they expect such system to comply not only to the confidentiality, authenticity and non repudiation guidelines given by World Bank related to Price quotation for e-procurement system, but also to comply with maintaining confidentiality of data related to Government buyer departments. The Security Model proposed here studies threats, vulnerabilities and risks to e-Procurement System, evaluates the suitability of Tokenization, Masking and Encryption techniques by applying them to ensure confidentiality, privacy and integrity of data related to bids, and security agencies. The study finds out that masking need to be applied to ensure confidentiality of data of security agencies, Public Key Infrastructure (PKI) to maintain confidentiality and authorization of price quotation and bid, e-sign to bind both buyers and suppliers to the actions taken by them in life cycle of e-Procurement process. Database security controls need also to be implemented so that data related to security Agencies and keys used to encrypt/ decrypt price quote are put in different tables/vault accessible only to authorized users excluding Data Base Administrator. Current approaches focus only on PKI.

KEYWORDS

Tokenization, Masking, Encryption, e-Procurement System, Privacy, Confidentiality

1. INTRODUCTION

Government e-Market (GeM) Place [1], a type of e-Procurement System, is a meeting place of suppliers and purchasers where purchase to pay process is electronically supported and a Government department defines sets of rules for procurement process. E-marketplace is a Business to Business relationship model (B2B) wherein multiple buyers can select products and Services from pre-sourced catalogues and perform commercial transactions with multiple sellers through a Web platform. The B2B model allowing direct purchases, procurement through bidding and reverse auctioning helps organizations in saving cost and increasing productivity enormously [2], so realizing the benefits of Government e-Market Place, several security agencies are also keen to join it. But they are concerned about the confidentiality and privacy of their sensitive data, the way suppliers were worried about the confidentiality of their price quotation before bid opening time. World Bank [3] issued the guidelines for it and Public Key Infrastructure based technique was used to address it. Similar solution for preserving confidentiality of sensitive data of Security Agencies is studied and being implemented through data masking technique. According to NIST [4], the growth of business, buyers and suppliers on GeM [1] depends upon its secure and reliable functioning. Studied
by Abdullah et al. [5] found, that security breaches result in loss of reputation, customers and economy of a country. As application design of e-Procurement System solution varies from one environment to another, the design of Indian Government e-Procurement System solution and Ketera used in UK and USA by one of the largest multinational companies in procurement of fuel and energy, studied by Juliette Stephens and Raul Valverde are considered to increase the surface to identify threats, vulnerabilities and associated risks and applicability of the proposed Security Model. This model is designed on the basis of cyber security research methodology, OCTAVE-S [6], It spans over four stages: 1) Assets Profiling, 2) Infrastructure Vulnerabilities, 3) Privacy and Confidentiality [7] Techniques and 4) Security Framework. Each stage of the security study is explained section wise in four subsequent sections of the paper and finally findings of the proposed security model are summarized at the end.

2. ASSETS PROFILING

The list of all assets such as servers, firewall, system software, etc. required for hosting e-Procurement System was prepared and assets were demanded from Cloud Service Provider (CSP). The owner Department of Government e-Procurement System checked that CSP produces evidence of implementing ISO27001 [8] Information Security Management System applicable controls in all concerned areas of Infrastructure and operation level. It was ensured that CSP is security certified and CSP has policies in place to prevent spread of viruses through computing infrastructure. It has a documented policy to apply system updates and patches to computing infrastructure provided by it and all the personal edge devices are updated, patched and have up to date antivirus protection.

3. INFRASTRUCTURE AND APPLICATION DESIGN VULNERABILITIES

Here vulnerabilities testing at System and Network, application source code and design, were conducted. Penetration testing tools were used to find the vulnerabilities at System Level, Network Level and application source code. Since tools cannot check vulnerabilities present in Application Design, application threats anticipated, threats identified in other e-Procurement Systems[9], Security incidents of 185 Japanese organizations studied by Abdullah et al. [5], and World Bank guidelines [3] were considered to construct Threat Model which is given as below:

- Actions of Anonymous and Validated Users (especially those performing Reporting and Analytics) were reviewed and checked if it was possible to see:
  - the products,
  - delivery locations and
  - Buyers’ details of security agencies in purchase orders.
- Actions of Validated Users in defined roles were examined and checked if it was possible to:
  - access or modify the information and privileges defined for other roles
  - access or modify information of other
Users defined in the same role
- attach fake budget approval or modify sanctioned amount or perform any
- other sensitive process of system anonymously
- attach fake inspection report and release the payment anonymously

• Examined, if it was possible to ignore Informing (through email/SMS) some of the registered suppliers and service providers about bid notifications to avoid one bid situation and get competitive bid price
• Verified the price quotations received are encrypted and kept in a Vault in a separate table on a separate database server, accessible only to buyer departments before bid opening time
• Verified if there exists any mechanism to bind the suppliers and service providers with their offers including discounts, if any.
• Verified if it was possible to tamper price quotations once opened

Success in any of the action as mentioned above results in compromise in the integrity or confidentiality of the application or back-end data. It will impact the one or the other stakeholders of the e-procurement system as shown in Table 1 and they may lose trust in e-procurement system.

Table I: Threat Assessment Model

<table>
<thead>
<tr>
<th>Risk</th>
<th>Severity</th>
<th>Actor Impacted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ability to see data of security agencies (buyers) related to orders, product procured and delivery locations and sharing with adversaries</td>
<td>High</td>
<td>Security Agencies</td>
</tr>
<tr>
<td>Ability to see price quote of suppliers before opening date and time of bid as it is not encrypted and disclose or modify it</td>
<td>High</td>
<td>Suppliers</td>
</tr>
<tr>
<td>Price quotes are in the same table where bid generated and accessible to DBA</td>
<td>High</td>
<td>Suppliers</td>
</tr>
<tr>
<td>Common Data Architecture shared between security agencies (e.g. same tables or different tables in same Data Base) and other Government buyers</td>
<td>High</td>
<td>Securely Agencies</td>
</tr>
<tr>
<td>Weak Authorization security control with no binding as to who created in following modules: Order Modules, Billing Modules, Inspection Modules, Payment Modules</td>
<td>High</td>
<td>Security and Other Agencies</td>
</tr>
<tr>
<td>No intimation of bids/Reverse Auctions sent to Suppliers/Service Providers when bid/reverse auction initiated</td>
<td>High</td>
<td>Security and Other Agencies</td>
</tr>
</tbody>
</table>

4. PRIVACY AND CONFIDENTIALITY TECHNIQUES

At this stage, the documented processes of e-Procurement System owner, articulating policy and procedure were examined. It was found that policy for following processes were present:
• Registration of Suppliers and Government buyer departments to ensure neither Supplier nor Buyer can be registered without proper e-verification and they cannot register multiple times to avoid submission of three bids by same supplier.

• User accounts and passwords to avoid adversaries to guess user credentials and access system

• Reporting security incidents so that if any event real or suspicious about information compromise comes in notice, it must be reported to Security Group.

It was found that the e-Procurement System owning department needs a policy to handle sensitive data of security agencies to build trust. So a study of three techniques: Tokenization, Masking and Encryption was undertaken to find out the most suitable among them to preserve confidentiality and privacy of the sensitive data.

4.1 Data Hiding Techniques

Security agencies also come on e-Procurement System portal to procure sensitive products such as weapons, uniforms, bullet proof vests, etc. and at the same time they want to maintain confidentiality about such procurements so that adversaries don’t know about details of delivery locations and indentors. They want their sensitive data to remain confidential, like:

• Identifier: Attributes that can directly and uniquely identify an individual, such as name, ID and mobile number.

• Quasi-identifier: Attributes that can be linked with external data to re-identify individual records, such as gender, age and area code.

• Sensitive Attribute: Attributes that an individual wants to conceal, such as department name, products ordered and delivery locations.

• Non-sensitive Attribute: Attributes other than Identifier, Quasi-identifier and Sensitive Attribute.

It can be achieved by applying the concept of k-anonymity [10] it is shown in Table 2 where sensitive information is concealed by applying concept of 6-anonymity where information of six attributes is concealed. 6-Anonymity With Respect To Department(Dept.), Order No., Product, Delivery Location and Email is shown in Table 2.
Microsoft [11], has given the concept of Differential Privacy where a Privacy Guard sits between user and the database server, it hides the sensitive information. Using techniques of k-anonymity.

The key question is which technique to choose Tokenization or Masking.

**a) Tokenization**

Tokenization [12] is a way, to replace sensitive data with non sensitive placeholders called tokens. The sensitive data is replaced with tokens in relational databases and files. The tokens are random values which replace original data but have no intrinsic value. For example, in a casino, money is replaced with tokens which is used for making payments and whatever number of tokens are left at the end, they can be converted to money in a controlled environment. A token used in place of cash cannot be used for doing financial transactions.

On the other hand, encryption is a way of protecting data by scrambling it into an unreadable form. It can be converted into readable form using right key. So attackers try to catch hold of encryption keys since with keys original data can be created. Compared to encryption, Tokenization is not scalable and suitable for unstructured data such as large files.

**b) Masking**

The general principal to preserve privacy is to suppress the sensitive data before it is subjected to disclosure. The sensitive data in the database is hidden. The identifiers which uniquely zero downs someone is suppressed, it is called as de-identification. One such technique of de-identification privacy is k-anonymity [10], A personal record is said k-anonymous if every record is distinct from at least k-1 other records over given “quasi-identifier” subsets of attributes. A quasi-identifier attribute such as age, gender and address, is one which gives clue about the identity of a personal record if linked to external dataset available with attacker.

Therefore, quasi-identifier attributes are completely or partially suppressed so that k-anonymity is attained. There are four ways to replace exact attribute value with less informative value, first replace the value such as 27.23 to 25.0; second do top coding, replace age above 60 to 60 and qualify it a senior Citizen, third do generalization, define address though...
zip code, fourth use intervals, define age 18 to 15-20, name Sharma to ‘S-T’, but if the requirement is for deep analysis of sensitive information such techniques cannot be used. Oracle [13], uses four masking techniques: 1) Condition based masking which is applied when two types of formats (e.g. pure numeric in one state and alphanumeric in another state) are used for storing same information, 2) Deterministic masking which assigns same masking value to a Social Security Number across all databases, 3) Compound masking is applied to related columns as a group (e.g. name is stored as first name, last name) to retain same relation and 4) Key based reversible masking where data is masked and recovered in original format using same key.

Masking Performance: According to the benchmarking done by Oracle [13], on an Exadata X2 - 2 full racks with high performance discs, 2 x 6 core Intel Xeon X5675 Processors (3.07 GHz). 1 TB data base containing a table of 6 billion rows whose one column was replaced with random numbers took 12 hours 49 seconds. 100TB data base containing a table of 600 billion rows where one column was replaced with fixed number took 33 minutes.

c) Encryption

Functional encryption is used to encrypt data; its aim is to ensure that an untrusted process learns only the output of a function f(x) about data and nothing else. Earlier engines were not allowing performing SQL operations on encrypted data, but work on CryptDB [14], has shown different SQL operations can be performed with 21-26% reduction in through put. CryptDB [14], sits between users and data base management system (DBMS) as a security guard. It uses different types of encryption algorithm for different types of columns depending upon the SQL operations performed on them. For example, if addition on two encrypted columns is to be performed, they are encrypted using Paillier [15][, cryptosystem since it allows addition on encrypted data. Similarly if multiplication is to be performed on two columns, multiplicative homomorphic encryption ElGamal [16] cryptosystem is used to encrypt them. For columns on which equality comparison, equi join and GROUP BY are performed such columns are encrypted by applying deterministic (DET) encryption algorithm.

Similarly if on a column ORDER BY or MAX or MIN SQL operations are to be performed then Order-preserving encryption (OPE) of Boldyreva[17] is used to encrypt such column. When a column is encrypted with OPE algorithm of Boldyreva [17], the server performs range queries giving encrypted constants OPEK(c1) and OPEK(c2) to DBMS corresponding to the range [c1, c2]. Using this encryption method, ORDER BY, MIN, MAX, SORT, etc. can also be performed. The randomized (RAN) encryption does not support any operators, and intuitively, the most secure encryption scheme is chosen for identifiers. It is implemented using Blowfish,[18] to encrypt integer values, taking advantage of its smaller 64-bit block size, and AES [19], encryption algorithm to encrypt everything else with a 128-bit block size.

Throughput of CryptDB [14] compared to MySQL is 21-26% slower and it is impacted maximum by SUM clause, which when performed on encrypted data brings it down 2.0 times less and when used in UPDATE Statement brings it down by 1.6 times. According to
Stephen Tu et al.[20], the disk space requirement of CryptDB [14] is 4.21 times more to the disk space consumed by plain text. If plaintext takes 17.10GB then CryptDB [14], takes 71.98GB Disk space.

5. CONSTRUCTION OF SECURITY MODEL

A security Model is a framework, which refers to a set of well-documented processes defining policies and procedures; and management of security controls of an Information Technology System. A security risk [21] is the loss potential to an asset of organization if at threat can exploit a vulnerability. The goal of security controls is to contain the security risk and protect information systems, maintain data confidentiality, availability and integrity of business processes. As threats and vulnerabilities vary for each application NIST [4], this Security Model for e-Procurement System is built by taking into consideration the security controls for the threats and vulnerabilities as identified in Table 1 above. These security controls are as follows:

I. Security Control for Risk 1 - Data Hiding for Security Agencies

Based on the study of above three techniques, it was found that Tokenization is not scalable and overheads of masking are less than the encryption. Further, masking allows SQL and mathematical operations on masked columns so the data of Security Agencies should be masked. It provides effective safeguard of data when it is at rest and also when used for preparing Reports and data analytics. The Secure Socket Layer(SSL) is used when data is on move.

II. Security Control for Risk 2 - Maintaining Confidentiality of Price Quote

The Price Quotations received from bidders are invited in encrypted format using Public Key Infrastructure. Either dongle can be used where Supplier encrypts the price quotation using Buyers Public Key and two employees of Buyer Department decrypt it using their Private Key in the presence of Suppliers at bid opening time and all of them e-sign them to make them tamper proof. Or all the price quotations are encrypted by applying RAN encryption algorithm and master key is stored in a separate Table called as Vault on a separate database server. At the bid opening time, the employee of e-Procurement System logins and the employee of buyer execute the process of decrypting price quote after feeding in One Time Password which the employee receives on his/her registered mobile number in the presence of Suppliers.

III. Security Control for Risk 3 and Risk 4 - Maintaining Keys and Data of Security Agencies in separate Tables

All the security controls are at Application Level so a long access control list is maintained. The application design is reengineered to strengthen the security by adding two database level controls, 1) data of Security and other agencies is separated out by storing them in different databases/tables and 2) as said earlier the master key used for encrypting price quote is kept in a separate table on a different database server accessible only
to e-Procurement System owner and buyers.

IV. Security Control for Risk 5-
Binding users to critical actions
through e-Sign Based Controls

Besides data, for sensitive processes performed by buyers such as entering budgetary approvals, inviting bids, placing orders, accepting consignment and inspecting them, releasing payments; and similarly for sensitive processes such as submitting bids, bills performed by suppliers e-Sign is made mandatory since it brings accountability. Both database level control and e-Sign based control as mentioned for Risks 3 to 5, lower the complexity which otherwise come in application code if they are implemented at application level control level viz., Access Control List.

V. Security Control for Risk 6-
Notifications of bids/Reverse
Auctions to Suppliers/Service
Providers through E-Mail and
SMS

Whenever a bid is floated by buyers on e-Procurement System, sending e-mail and SMS are made compulsory to registered suppliers to avoid vendor cartel, biasedness, get competitive bid price and bring transparency. If such intimations are not sent due to one or other failure, it can invite trouble so auditing of actions performed by buyers, suppliers and e-Procurement System owner is made mandatory.

6. IMPLEMENTATION

Based on the information provided in a study conducted for Asian Development Bank [22] for six e-Government Procurement Systems; and information collected for Government e-market Place through interview and code inspection. The key focus of all systems is on keeping confidentiality of bids. The systems can be divided into two groups. There is no information about keeping confidentiality of sensitive information of buyers.

The security framework is applied on systems of two procurement Groups “A” and “B” at four layers: Infrastructure, Application, data, and Process and their security posture is measured.

I. Infrastructure Layer: The entire infrastructure needs to be secured at the perimeter level so Firewalls and Intrusion Prevention System must be installed and network segmentation should be done. Also hardening of Operating Systems of Routers, Servers and management Workstations should be done.

Findings: Infrastructure of both Groups is secured.

II. Application Layer: The Application design assessed to ensure that it protects the confidentiality, authenticity or integrity ISO 27001 [8] of bid information and other sensitive data of buyers. Here PKI based asymmetric cryptographic Controls and Masking need to be used to encrypt and hide the data. Key management vault shall be
stored separately from encrypted data to support use of cryptographic techniques and DBA should have no access to it. E-sign should be based on two factors authentication [23], the second factor could be one time password or biometric etc. to prevent others to impersonate and sign on a document. The authentication mechanism should ensure credentials are given on a page, which is under SSL.

**Findings:** The bid is encrypted at Client end in systems of Group “A” and sent through SSL enabled page of client, while in second system “B” bid is encrypted at server side and sent through SSL enabled page of client.

- System “A” has separate vault to store Private Key of Buyer Officers opening the bid, while System “B” stores them on same server where encrypted bid is stored.
- System “A” uses hardware based dongle for Public-Private Key pair whereas System “B” uses e-Sign which uses Social Security Number for signing a document along with OTP.
- E-sign cannot encrypt/decrypt documents as it is not its mandate but dongle can do both bid encryption/decryption at client side to achieve confidentiality and e-sign to achieve integrity and non repudiation.
- So system “B” uses encryption/decryption of bid Price at Server side through RDBMS supported encryption algorithm.

### III. **Data Layer:**

Sensitive data should be encrypted, masked or hashed in the database.

- Application design should differentiate between data that is sensitive to disclose and must be encrypted using PKI asymmetric key (discussed above).
- Data (bid price after opening the bid) that is sensitive only to tampering must be e-signed and a keyed hash value (HMAC) must be generated.
- The data that is revealing identity of a security agency should be masked to preserve its privacy.
- The data that can be irreversibly transformed (hashed) without loss of functionality (such as passwords) should be treated separately.
- Access control method at Application level should be enforced to provide access to sensitive data and functionality only to permitted users or clients.
- Role-based access controls should be enforced at both database level and application interface. This will protect the database in the event that the client application is exploited.

**Findings:** It is an important finding so far little attention is paid toward this since it is a post bid opening data.

- Both systems are keeping buyers’ information in plain text, which they need to mask.
- Both are using common data architecture for storing information of security agencies and other agencies.
- Both systems are not keeping e-signed bid data.
IV. Process Layer: At least following IT security processes and best practices need to be in place for operation and maintenance of Procurement System in line with international standard on Information Security Management System, ISO 27001 [8]

Findings: System “B” disaster recovery site is being designed; it is not available right now. Disaster recovery site must be in place to ensure availability and audit log shipping. As logs are not written for spyware planted at kernel level so logs should not be treated as the sole protection against mala fide acts. Anti spyware, anti spam and antivirus software should be installed. Other findings for both systems are common, which are as follows:

• The application hardening done for Top 10 vulnerabilities defined by OWASP [24]

• Network security assessment done for adequate security through penetration testing and vulnerability assessment as per NIST SP 800-115 [25].

• The software source code evaluated using white box test approach through code review/inspection process for identifying malicious codes/ Trojan etc.

• End to End transaction workflow checked to verify it is going through the defined path by using dummy test transactions

Keat et. al. [26] conducted a study in Malaysia to know confidence level of people about keeping the data on cloud storage. It revealed that people feel that investigators have limited skill set, limited jurisdiction under their authority and there is shortage of forensic experts so they would not be able to investigate criminal incidents. Further the laws, regulation and guidelines are also inadequate so these will not be able to provide advisory at the time of need. Thus GeM system is hosted on a Private Cloud maintained by a Government agency.

7. CONCLUSION

The proposed Security Model extends existing security foundations of e-Procurement System by adding four security controls at 1) Database layer, 2) Masking and Encryption layer, 3) e-Sign layer and 4) Alert layer. The inclusion of Masking which hides sensitive data but allows SQL and other operations continue to be performed on it, is the hallmark of the proposed Security Model. When masking is compared with Tokenization, it is found Tokenization is not scalable for large number of records. Although encryption as mentioned in CryptDB [14] has come a long way where it is possible to apply different types of algorithms on different columns depending the purpose for which they are used in the application and carry out different types of operations on encrypted columns. Downside is management of encryption keys of different columns, overheads related to extra disk space requirements and degradation of CPU response time. But still, it works better than Tokenization.
For hiding data of buyers, masking technology is better than both Tokenization and Encryption. Masking, similar to encryption scrambles the data without disturbing the format by working on full database or subset of dataset. The aggregation, preservation of sum and average values over masked values can be done. Masking can be performed in such a way that it’s extremely difficult to reverse engineer the original values. It helps in guaranteed security of data.

REFERENCE

3. World Bank : Guidelines Procurement under IDRB Loans and IDA Credits, available on t