Winston Shine

ILC Database Theory/Security

Week 3 Progress Report

2/16/2023

My goal is to study Database Theories and Security through the book Translucent Databases. Some additional smaller goals I am trying to accomplish are doing my work this quarter with docker containers and If time allows study a bit about document-oriented databases.

This week I learned about applying techniques to a different data structures, namely XML. A lot of the reading was going over something known as the XML Security standard. Three of the sub categories of the XML Security standard are Canonical XML, XML Signature, and XML encryption. Canonical XML is a standard that is used to force XML files to have a similar form. XML files can use tags in separate ways or have attributes listed in different orders. For example if two xml files representing the same data listed the attributes in a different order, depending on how one way functions were being applied this could lead to different outputs for both files. Another big topic in XML canonicalization is how is white space treated. From what I understand is there are libraries used that provide algorithms to canonicalize an xml to ensure it meets the XML Security Standards before it is used for encryption or signing.

An xml signature loos like a list of information under a <signature> tag added to the end of the file , but still under the root element. Some of the information listed in this signature is : CanonicalizationMethod, SignatureMethod, Reference URL, Digest Method, Signature Value, X509 Certificate, and Key information.

When encrypting XML files, any nodes are encrypted and the data inside is replaced with encrypted information, the tags are not encrypted. Under an encrypted node several new nodes are used to specify information about the encryption, in one example these were EncryptedData, EncryptionMethod, CipherValue, Cipher Data, Key info, and Encrypted Key. I think the downside to using XML to store encrypted information would be the extra space needed when storing a large amount of data. To encrypt one node numerous new nodes need to be added to describe the encryption done.

I also started reading about Quantization and although I plan to finish reading about it in the following week I will try to explain what I understand so far. Quantization can be described as separating data into groups, normally called quanta. The key here is that the data is being rounded off so we actually lose information. An example the book used that I found helpful was that “Rounding real numbers is equivalent to creating quanta I - .5 <= x < I + .5 for all integers I.” (p. 61 Translucent Databases). From the examples given it seems quantization is used to obscure or even misdirect someone to what the data actually is. Some of those examples are: ‘six figure contract’ replacing an actual number which could be in the range 100,000-999,999. Or GPS data containing location date and time, might leave out portions of the time to obscure the data. This last example is a technique of quantization is called Deleting Smaller Fields. One of the techniques I found interesting since it seems opposite of the previous one is to Add random Amounts. This technique is to simply add random amounts of information to data, which doesn’t make the data less accurate or hide anything but makes it harder to make sense of.

For Next week I plan to finish reading about Quantization and then move on to Chapter 7 of Translucent Databases called Coordinating Users.