Fundamentals of System Security
(COMSM0122)
Lab 1

Sensitive Data Classification

Scenario

PiC is a new online pharmacy system that differentiates itself by promising the most data protection centric fulfilment service on the market. Patients are delighted with the prices and the convenience of the service. But that convenience and speed does not come for free. It takes access to customer data and it also takes APIs that can transform that data. Given the sensitivity of this data, PiC must implement best practice security and privacy engineering across board.

You've been tasked with implementing data protection throughout the system. Here is PiC's data flow. PiC has about 50,000 customers across the United Kingdom (see `customers.json`). Most customer interactions happens through their mobile app. PiC's existing operations stores customer data and some of that data is very sensitive. PiC also collects data about its employees, contractors, visitors, orders, inventory, deliveries and finances in seperate datastores. All collected data is operational and full of sensitive and precise information required to run the business. Currently, all employees have access to all data including `customer.json`. No data has been shared externally so far, but it will be soon as the PiC grows its customer base. Each unit within PiC uses `customer.json` to improve their understanding of the business, to improve performance, and in some cases to prove that PiC is compliant with data protection regulations.

PiC has a specific business requirement to perform certain analyses. These include predictive drug ordering; HR compliance surrounding topics like age, sex, and race; understanding the medication used at the city and county level; and delivery efficiency.

Objective

You will use above fictional description of PiC in the following hands-on lab. The objective is to enable PiC business requirements with maximum data protection balanced with the need for data to run the business. In this Lab, you will transform some of PiC's customer data to serve their business needs while protecting their customers. You will segment the data into separate datasores and recommend appropriate access control and retention policies. You will focus on the PiC's operations team need to manage inventory. The operations team currently use delivery data for inventory analysis. The data include order information like drug type and volume and personal information about each customer, include their chi number as prove of identity. While all this information is necessary to support deliveries, some of it is highly sensitive. Storing this data presents security and privacy risk. The aim is to apply your knowledge of data classification to reduce that risk.

Task 1:

Based on above scenario, review `customer.json` datastore (see snapshort in listing 1), consider how you will manage associated risks.

1. What is the appropriate classification for the datastore?
   Answer: This data contains sensitive information about customer data. It should be classified as restricted

2. What sort of access should be granted to PiC employees?
   Answer: Access should be highly controlled, limited and monitored.

3. How long should PiC store a customer's information in its datastore?
   Answer: The data should be kept for as little time as possible (ie only until orders are delivered and return/refund time expires)

```
[
  {
    "name": "Aaberg Zysk",
    "chino": "391-081-878",
    "medication": "Metoprolol Tartrate",
    "county": "Abertillery",
    "city": "Gwent",
    "ethnicity": "Jin",
    "volume": 10
  },
  ...
]
```

Listing 1: Snapshort of `customer.json`

Task 2:

This activity assumes a basic understanding of javascript and knowledge of how to iterate over `json` objects. There is a reasonable and expected tension between level of privacy, security and business goals in this scenario. Finding and maintaining an appropriate balance is the first priority for any secured system. Being overly sensitive or overly restrictive will inhibit the ability to analyse data over time, which is a key business imperative. Your task is to create a new way to manage the flow of data that will enable PiC to do the necessary analysis while minimising the risk to data privacy.

1. For this analysis you need only data points for medicine name, volume and general location. Generate a new datastore that only contains stated data points. Save the datastore as `medicationbyvolume.json`.
   Answer:

```
   const fs = require(’fs’);
const path = require(’path’);

async function generatedrugbyvolume() {
    const customerfile = ’customersdata/customers.json’
    var customers
    var medicationslist = []
    var medicationvolume = new Array()

    try {
        if (fs.existsSync(customerfile)) {
            let customerdata = fs.readFileSync(customerfile, { encoding:
                ’utf8’, flag: ’r’ });
            customers = await JSON.parse(‘${customerdata}‘);

            for await (let customer of customers) {
                medicationslist.push(customer.medication)
            }
            medicationslist = [...new Set(medicationslist)]

            for await (let medication of medicationslist) {
                let volume = 0
                for await (let customer of customers) {
                    if (‘${customer.medication}‘ === ‘${medication}‘) {
                        volume = volume + customer.volume
                    }
                }
                medicationvolume.push([medication, volume])
            }

            //save medication volume
            const medicationvolumedir = ‘${__dirname}/medicationvolume‘
            const medicationvolumefile =
                ‘${medicationvolumedir}/medicationbyvolume.json‘
            async function createdir() {
                if (!fs.existsSync(medicationvolumedir)) {
                    fs.mkdirSync(medicationvolumedir, {
                        recursive: true
                    });
                }
            }

            function replacer(key, value) {
                if (value instanceof Map) {
                    const obj = {}
                    for (let [k, v] of value)
                        obj[k] = v
                    return obj;
                } else {
                    return value;
                }
            }

            await createdir()
            let medicationvolumedata = JSON.stringify(medicationvolume,
                replacer, 2);
            fs.writeFileSync(medicationvolumefile,
                Buffer.from(medicationvolumedata));
        }
    }
    catch (err) {
        console.error(err)
    }
}
```

Listing 2: Snapshort of `customer.json`

2. Currently all employees use the `customer.json` datastore. With the new addition

of `medicationbyvolume.json`, how would you apply classification labels to achieve effective retention and access control policies necessary to secure the system.
Answer: Review PII/quasi attributes on data store to determine classification labels and level of access control.

3. Which of the following access design strategies would you recommend?

   (a) Remove all access from the operational deliveries database and give everyone access to the analysis databases.

   (b) Restrict access to the operational database to only the operations team, and give everyone else access to the analysis databases.

   (c) Restrict access to the operational database to only the operations team, and give access to the analysis databases to only those who need it

   (d) Continue to allow everyone to have access to everything

4. How can you improve PiC's data retention policies based on `medicationbyvolume.json` data objects?

5. Which of the following retention policies would you recommend?

   (a) Retain data in both operational and analysis databases indefinitely

   (b) Set the retention policy for data in analysis databases based on time required to complete analyses and allocated storage budget, and delete delivery data once order completion is confirmed

   (c) Let the analysis and operations teams delete data when they are concerned about data volume and/or storage costs.

   (d) Delete data entries in both operational database and analysis databases if the entry has not been accessed in the previous week

6. You are reviewing a database to evaluate privacy protections. The data contains sensitive information that could clearly identify individuals. Which of the following is true?

   (a) The data should be classified as ”confidential”

   (b) Retention should be extended to allow for analysis

   (c) Access should be highly controlled, limited, and monitored

   (d) All of the above

   (e) B and C only

Deliverables

This is a formative assessment. Ensure one of the TAs or the course lecturer has reviewed your solutions before the end of tutorial session.