BRITACOM Virtual Seminar 3, 8 April 2021

Tax-related Data Governance and Application

Philip Chew, Architect, Infocomm Division, IRAS

IRIN – Inland Revenue Interactive Network (Slide 3)

The System we use is called the Inland Revenue Interactive Network, or as we more commonly call it – IRIN. IRIN is used to administer taxes in Singapore.

1. We serve a range of external users such as individuals, companies and other organisations such as charities. We also work with other government agencies, foreign tax authorities, etc. where we exchange information or request for specific services.

2. We interact with taxpayers through various ways, including our website, call centre, front-line desk services and various correspondence methods like emails, digital notices or SMS. We use APIs and SFTP to perform transfers with other agencies.

3. Our internal users consist of tax officers, IT staff, temporary officers (such as those on short term contracts) and vendors. Our biggest group of users are our tax officers who use the internal interfaces of IRIN to perform their daily work and review documents submitted by our taxpayers.

4. Our tax officers have access to a range of systems such as document management, workflows, intranet services, reports, analytics platform, customer management system, etc.

5. Both External and Internal Systems share a common Core Processing framework, containing common data and processing services. This provides reusability and standardization of application design, infrastructure and security controls.

Enhancing our Capabilities (Slide 4)

Over the past few iterations of IRIN, we have already achieved some key capabilities that enable us to shorten processing time with increased efficiency. With the upcoming iteration of IRIN, we look to:

1. Move onto the Government Commercial Cloud, a special subscription where the Singapore Government has signed up with Cloud providers to create a more controlled and secure environment where government IT systems may be hosted, while retaining the flexibility and efficiency of Cloud services like PaaS and SaaS. At the same time IRAS is upgrading the way we develop systems, by incorporating security into the development pipeline to ensure our code and packages we use is kept secure from development to production.

2. We move into Microservices Architecture, where we decompose large complex services into smaller services with single responsibilities. This level of granularity helps us segregate our data, allowing us to apply security controls befitting the content of the data, while reducing the surface of attack because of large, combined datasets.

3. We leverage on data and design driven approaches such as identifying the types of data we want to collect and developing ways to use and model the data to further discover new insights.

4. With the move to the Cloud, we relooked at the way we do security to develop a zero-trust model and customized risk management measures, which will be covered in the upcoming slides.

Developing our Security Strategy (Slide 7)

1. Threats – Traditional methods of perimeter defences are no longer enough. Attackers have developed new techniques that attempt to compromise systems from within, such as advanced malware or compromising commercial software. We perform threat modelling based on the Assume Breach mindset to model the kinds of attacks we may encounter.

2. Enhance Capabilities – With IRIN 3 moving onto the Cloud, we have been progressively putting into place enhanced detection and response strategies while continuing with what we have done for preventive security. Our focus is on slowing down attacks and increasing our window of opportunity to detect and respond to any attacks.

3. Develop Solutions – We implement zero-trust, a model where we trust nothing, regardless from within or without the system to mitigate the threats we have identified. The model Trust nothing, verify everything.

The Zero Trust Model (Slide 8)

With the Zero Trust Model, we focus on 3 key areas

1. Data – All databases are encrypted in storage using keys provided by IRAS (also known as bringing your own key), and in transit as the data moves across the system. We use data tokenisation to protect PII fields as it moves between systems. The end-state of our data protection includes setting rules on queries and monitoring those queries using Database Activity Monitoring.

2. Infrastructure – We place an emphasis on Governance and Risk Management to make them suitable for the Cloud. We will be implementing a full-fledged Security Operations Centre or SOC that will be responsible for monitoring and responding to security incidents.

3. Identity – Last but not least, Identity forms the 3rd part of a good zero-trust structure. Identity includes not only details who the users are, what they can do and should also include how they are accessing data and where they are doing it from. All user and administration activities are tracked and logged on the Cloud as well.

Setting Up the Overall Strategy (Slide 9)

1. One of the first things to protecting a system would be to set up an overall strategy to guide the sort of protection a system requires. We need to identify the “crown jewels” of a system, for example, key databases holding highly sensitive data such as taxpayer information. A Risk Management Framework is used to identify threats, and how such threats can be mitigated. The governance framework is set up to develop guidelines and policies to based on the risks identified.

2. The outcome should be a matrix showing how a threat category is being mapped to a security control, and how the security control has been implemented in the System. The following shows a series of examples of threat to control mapping. Shown here are some of the steps done during an attack, such as reconnaissance to identify potential loopholes, delivery of an attack and what is done during an actual attack.

Identifying Users, Controlling Access (Slide 10)

1. As mentioned earlier, we need to know who is accessing the data, why they need access to the data and how they are accessing the data. User access control therefore becomes a key part of data protection measures. User account governance and management should be managed centrally with properly defined workflows.

2. Fine-grained access control can be done to break down users into groups or individuals, with special access to certain types of data.

3. Flow Diagram:
    a. We implement an identity governance solution to handle user creation, access requests, etc.
    b. The Identity Access Management solution is used to decide what kind of services or data users can access.
    c. A security token generated by the Identity Access Management is used to validate users in the microservices.
    d. The identity of the user checked again at the database to determine what queries can be made by the user.

Data Encryption & Tokenisation (Slide 11)

In order to protect the data in its physical form, we may consider encryption or tokenisation.

1. Encryption scrambles the data using a key into an unreadable format. Keys used for encryption are provided by IRAS and are stored securely. There are a few options to encrypt data, ranging from encrypting each data field individually to only encrypting at the storage level depending on the classification of the data or potential risks in the environment.

2. Tokenisation basically replaces the content of the data with another content to hide the original content. This is reversible. For example, an address 123 ABC Road may be replaced to 821 Q17 bXaZ. Without access to the tokenisation solution, an attacker will not be able to decipher the original data. One of the ways IRAS uses tokenisation is to hide personal identifiable information (PII) as the data is passed through external environments.

Pro-Active and Reactive Controls (Slide 12)

In general, there are 3 different ways we can use to protect data. Preventive, Pro-Active and Reactive Controls.
Preventive Controls – For example what we discussed earlier, like identity access control, database encryption, etc. can be set up ahead of time and are used to prevent or delay attacks as they happen.

1. The Database Access and query monitoring can be an example of a Pro-Active or Reactive Control, depending on how it is set up. Pro-Active controls seek to rectify a security situation based on the data available. A DAM monitors the way a user or application is querying a database and based on patterns it recognizes, automatically denies the user from accessing the data. A DAM may act as a reactive control, meaning it passively collects the data and feeds it to a monitoring solution.

2. The Data Loss Prevention or DLP is another form of proactive control, that scans endpoints (for example the machines used by our staff or servers we deploy) for sensitive data, and patterns of movement or usage of that data to ensure that no sensitive data is being leaked from the system.

3. The Endpoint Detection and Response is a form of proactive and reactive control that scans endpoints for malicious activities and aids in stopping the activity and rectifying the damage done.

Monitoring and Tracking (Slide 13)

IRAS has a Security Operations Centre that centralizes the logging and monitoring of all alerts raised by a system.

1. The SOC comprises of different roles, such as specialized team members designed to monitor and escalate alerts, hunt down threats by manually analyzing logs for anomalies or coordinate a response to a security event.

2. An SOC has a range of tools at its disposal to ensure the Security of an organization is not compromised. The SOC primarily performs security monitoring. All events and logs are consolidated within the SOC and scanned for any potential events that might be a Security Alert.

3. The SOC uses technologies such as the following:
    a. SIEM - Collects and correlates logs and events generated from applications, security devices and systems supporting the 3 applications, alerts potential security threats based on predefined rules.
    b. SOAR - Automation and orchestration of alert responses, threat hunting activities and incident responses, Reduces human intervention with predefined workflows and automated responses
    c. UEBA - Aggregates logs and events generated from applications and uses machine learning to baseline normal user behaviors, apply analytics to identify risky deviations and alert suspicious behavior with high risk scores.

INLAND REVENUE
AUTHORITY
OF SINGAPORE

The information presented in the slides aims to provide a better general understanding of taxpayers' tax obligations and is not intended to comprehensively address all possible tax issues that may arise. This information is correct as at the date of presentation. While every effort has been made to ensure that this information is consistent with existing law and practice, should there be any changes, IRAS reserves the right to vary its position accordingly.