University of Auckland | ASB Bank

BTECH 451

Project Report

Implementation of Behaviour Profiling and Threat Detection Techniques in a SIEM Environment

Muhammed Raafey Ali **KHAN**

Department of Computer Science

University of Auckland

Auckland, New Zealand

mkha411@aucklanduni.ac.nz

Abstract

This report outlines the progress of a Bachelors of Technology (Honours) final year project being conducted at the University of Auckland. My project is to evolve, grow and mature the security practices currently being implemented at ASB Bank through the use of the Security Information and Event Management(SIEM) platform. In this report I outline the full years work, including a case study of Carbanak, internal research of ASB's SIEM Implementation, a literature survey of current behaviour profiling methodologies and an implementation of two systems within the SIEM platform. These two systems include an Information Security Overview, which presents all threat related activity in the ASB enterprise and the Behavioural Profiling System that looks profiles the Executive Leadership Group's user behaviour in order to detect potential compromises and misuses of these accounts.

Contents

1 Acknowledgements

Dr. Aniket Mahanti

Academic Supervisor

Senior Lecturer

University of Auckland

Auckland, New Zealand

a.mahanti@auckland.ac.nz

Dr. Sathiamoorthy Manoharan

B.Tech Co-ordinator

Senior Lecturer

University of Auckland

Auckland, New Zealand

s.manoharan@auckland.ac.nz

Malcolm Allen

Industry Supervisor

Information Security Specialist

ASB Bank

Auckland, New Zealand

malcolm.allen@asb.co.nz

Ryan Cotterell

Industry Supervisor

Head of Information Security

ASB Bank

Auckland, New Zealand

ryan.cotterell@asb.co.nz

And a special thanks to the entire Information Security Team at ASB Bank.

Disclaimer: *Please note, due to the nature of the data, several of the figures have sections blurred/blanked out. This was done intentionally to protect the confidential property of ASB Bank.*

2 Project Introduction

2.1 The Company

ASB Bank was first established in 1847, known at the time as Auckland Savings Bank and is now owned by Commonwealth Bank of Australia. ASB provides a range of financial services such as retail, rural and business banking, as well insurance services through its subsidiary Sovereign and investment and securities services through ASB Securities. ASB employs over 5,000 people across New Zealand and is considered a leader in technology innovation [15]. The bank has had success by being the first to introduce innovative services and features to the people of New Zealand, including:

1997 – Internet banking

1998 – Open branches seven days a week

1999 – Online share trading on both New Zealand and Australian markets

1999 – Mobile banking via ASB Mobile

2003 – Ability for customers to opt-out of paper statements

2003 – Introduce 2-factor authentication for Internet Banking (NetCode)

2006 – PDA and browser based banking

2012 – ASB payments via Facebook (iOS, Android and Windows phone)

2012 – Dedicated retail Application

ASB also employs a full-time information security team as well as a full time operational security team. In my time at ASB, I have had the opportunity to work with and learn from members of these teams in order to achieve the objectives of this project.

2.2 Project Brief

The overall goal of this project is to evolve, mature and grow the security practices currently being implemented at ASB Bank. ASB is continually looking to mature and raise the level of efficiency and effectiveness of its Security Information and Event Management (SIEM) implementation and associated practices. This project is being used as an opportunity to evolve and mature existing practices regarding threat profiling and detection and the use of behavioural profiling to improve on the detection.

The project is broken into two phases:

1. Threat Overview System
    - Establish a customized system that provides a single view of potential threats
    - The single view presented is to be based on defined pattern rules
    - The defined pattern rules are to be established through research and analysis
    - All aspects of the dashboard are to be automated, scalable and changeable
2. Behavioural Profiling System
    - Devise an easily adoptable and scalable behavioural profiling and base lining threat detection scheme
    - Deliver a working prototype of the above-mentioned system
    - As part of the recommendation include a working implementation of this methodology
    - The working implementation will be based around the system administrator user community

I have considered multiple aspects when designing the system including resource and process requirements, platform limitations, and complementary practices. Although a brief was specified, there was no set criteria given. ASB allowed me the freedom to use my creativity to come up with a viable solution.

3 Project Motivations

3.1 Cyber Crime Landscape

The cyber crime landscape is constantly changing and evolving. It is important for organizations, especially financial institutions to evolve and counter this threat. McAfee [16] estimates that the cost of cyber crime to the global economy could be as high $575 billion per year. This takes into account both the gains to criminal organizations and the costs to companies for recovery and defence.

Ponemon Institute [12] conducted a worldwide survey of 257 companies across seven countries to assess the state of cyber crime in the world. Some of the results they were able to gather from this survey were:

- Cyber crime continues to be on the rise for organization, there was a 10% increase in the annualized cost from 2013-2014
- The cost to cyber crime varies depending on organization size. There is a positive relation between organization size and annualized cost
- There is a positive relationship between the time to contain an attack and the cost to an organization. If attacks are not resolved quickly, it can get very costly for a company. The average time to contain a cyber crime was 31 days, with an average cost of approximately US$640,000, representing a 23% rise from the year 2013 to 2014.

The way in which cyber criminals are attacking targets has now changed; they are no longer rogue hackers, but organized groups. Because of this, it is important to detect a threat early. Through this project I aimed to improve upon the security practices at ASB, with the objective to empower enable security professionals to detect and assess potential threats in their early stages. An improved and streamlined threat detection scheme would reduce false alarms and enable the security team to focus on real threats, thus saving time and money.

This project deals with creating automated threat visualization systems that provides a broad overview of the security concerns in a given time frame. This will allow the user to spend more time looking at potentially suspicious activity and make better decisions related to management of time and resources.

3.2 Carbanak Case Study

To fully understand the ramifications of targeted attacks on systems, it is important to look at particular instances where enterprise systems have been compromised. For this report I will be looking into one particular such occurrence, Carbanak.

I will be using the cyber kill chain model as presented by from Lockheed Martin [8] to evaluate the case of Carbanak. The definitions for these kill chain phases are as follows:

Reconnaissance Research and identification of targets. Often done by the means of social networks, social engineering or browsing the web for conference information etc.

Weaponization Using a control action Trojan in combination with a deliverable payload. Usually done by the means of an automated tool known as a weaponizer.

Delivery Transmitting the weapon into the target system. Generally done by the means of mail attachments, websites and USB removable media.

Exploitation After the weapon is delivered to the host, exploitation code is triggered. This code may look to take advantage of system software or the OS itself.

Installation Installing a remote access Trojan inside the compromised system, allowing the attacker to stay persistent.

Command and Control (C&C) A compromised host must "beacon" to an internet controller server to establish a channel. Once this is complete the attacker may have "hands on the keyboard access".

Action on Objectives The attacker may now take actions to achieve their objective. This in most cases is data ex-filtration. It is also possible that the attacker only wanted to gain access to the fist compromised machine so they may move laterally though the system and compromise other machines.

Carbanak is the name given to a series of attacks on financial institutes around the world. The factor which sets Carbanak apart from other such attacks is that the attackers did not see data, but money as their primary target. The attackers were able to extract between 2.5 - 10 million US Dollars each from financial institutes and the total approximate loss is measured to be at around 500 million - 1 billion US Dollars based on different reports [7]

From late 2013 onwards, banks and financial institutes across the world have been attacked by a group of unknown cyber criminals. These attacks continue to this day with the attackers not having been identified.

This case study will be a technical analysis of these attacks using the cyber kill chain model. Kaspersky Lab HQ [7] has released an extensive report into the proceedings and the information in this section is taken from said report.

3.2.1 Reconnaissance

The reconnaissance phase for the Carbanak attacks was particularly complex and at times lasted for months before the attack actually occurred. It has been discovered that video recordings of bank employees, particularly system administrators were taken and sent back to the C2 server, meaning the reconnaissance phase continued on through the entire cyber kill chain process.

Once access is achieved to the infected system, attackers perform manual reconnaissance of the victims network to find vulnerabilities that may be exploited. Based on the results of this reconnaissance, more machines are compromised to find the correct system or user that will be used to conduct the attack. Sort of like target identification

3.2.2 Weaponization, Delivery and Exploitation

All of the cases observed so far have used spear phishing emails with vulnerable versions of Microsoft Word 1997 - 2003 attachments. The attachments exploit both Microsoft Office and Microsoft Word. The exact CVE numbers are as follows:

- Microsoft Office CVE-2012-0158, CVSS Score: 9.3
- Microsoft Office CVE-2013-3906, CVSS Score 9.3
- Microsoft Word CVE-2014-1761, CVSS Score 9.3

The recipients of these phishing emails were all employees of the target institution. The emails appeared legitimate and were at times sent from a co-workers compromised account, using compromised systems as a transmission vector.

The victim institutions were mostly Russian; therefore the names of the infected attachments found have been mostly in Russian. Examples of names include -115 which roughly translates to "Accordance with Federal Law", this is in most cases enough to induce an employee to click on and open the attachment. The following is a translated version of a Carbanak phishing email:

Good Day!
I send you our contact details
The amount of deposit 32 million rubles and 00 kopnecks, for a perios of 366 days, % year—end contribution term
Sincerely, Sergey Kuznetsov;
+ 7 (953) 3413178
f205f@mail.ru

In the case of this spear phishing email, the attachment was a compressed Roshal Archive (.rar) file.

In addition to spear phishing, it is also possible that exploit kits were used in conjunction to perform drive-by-download attacks, a malware delivery technique where data is downloaded by the target machine by simply visiting a website. Traces of the Null and RedKit exploit kits have been found on compromised systems

Once the user has clicked and/or opened the attached file, the remote execution vulnerability is exploited. This installs Carbanak onto the victim system.

3.2.3 Installation

Once the exploit in the phishing email or exploit kit is able to execute its payload, Carbanak is installed onto the victims system. Carbanak will initially copy itself into "$\%system32\%\backslash com''$ with the file name of "svchost.exe". It also has the file attributes of hidden and read-only. At this point the original file which is created by the exploit payload is permanently deleted.

Carbanak will in the next phase ensure that it has auto run privileges by creating a new service. A naming convention is in place to ensure that the given name does not appear suspicious to the average user. Before creating this service, however a number of tasks are performed in order to bypass Internet security and anti-virus detection, this is different for every organization attacked and comes from the reconnaissance phase of the attack.

Carbanak then creates a file with a randomly generated name and a .bin extension in
$\%COMMON\_APPDATA\%\backslash Mozilla$
where commands are stored that are executed later on in the process.

The next step in the process is to get the proxy configuration from the registry entry:

$$[HKCU\backslash Software\backslash Microsoft\backslash Windows\backslash CurrentVersion\backslash InternetSettings]$$

And the Mozilla Firefox configurations file in:

$$\%AppData\%\backslash Mozilla\backslash Firefox\backslash < ProfileName > \backslash prefs.js$$

Carbanak injects its code into svchost.exe and is now able to communicate with the C2 servers

3.3 Command And Control (C&C)

At this stage, the system is infected with the Carbanak malware, which is able to communicate back to its C2 server. Carbanak now logs keystrokes and takes screen shots on a regular basis, typically 20 seconds.

To communicate with the C2 server, Carbanak will use the HTTP protocol with RC2 +Base64 encryption. It will also insert strings with extensions such as .gif, .pdf etc. at random locations in the HTTP request.

Carbanak will send the collected data to the C2 server and also receive commands. The incoming commands are compared with a hash table. If a match occurs, the associated action is performed.

3.4 Action on Objectives

After fully surveying the victim organization for a prolonged period of time, the attackers were able to use intelligence gained from video and other monitoring techniques to create a profile of the typical user. This allowed the attackers to personalize each attack to the target organization. Examples of malicious operations conducted include:

- Creation of fake transactions in internal databases after the verification process, therefore avoiding discovery of fraudulent activity.
- Use of internal command utilities to insert fraudulent operations into the transaction queue
- Control over the internal ATM Network, If the bank had enabled remote access to ATM's, the criminals used standard utilities used to control and test ATMs to dispense cash at their will
- Retrieve sensitive bank documents such as emails, manuals, passwords and crypto keys. A particular example of this being a document found on a Carbanak C2 server outlining ATM keys used to verify the integrity of ATM pins being entered.

4 The Platform

In this section of the report, I will be outlining the platform I will be using as well as some of the key security mechanisms that have been put in place by ASB in regards to the SIEM platform.

4.1 SIEM

SIEM is seen as the integration of two key technologies, Security Information Management (SIM) and Security Event Management (SEM). SEM deals with real time monitoring of data, event correlation, notifications and console view, whilst SIM deals with the long term storage, analysis and reporting. For this project, I will be mainly working with the McAfee SIEM platform that is currently being implemented by ASB Bank. At ASB the current SIEM implementation was first introduced in 2014, taking over from its predecessor Nitro. SIEM is used to gather, correlate and report on data within the ASB enterprise.

The key features we see in most SIEM/Log Management solutions of SIEM systems include:

Log Aggregation Collection and aggregation of log records from the network, security, servers, databases, identity systems, and applications.

Correlation Attack identification by analysing multiple data sets from multiple devices to identify patterns not obvious when looking at only one data source.

Alerting Defining rules and thresholds to display console alerts based on customer-defined prioritization of risk and/or asset value.

Dashboards Presentation of key security indicators within an interface to identify problem areas and facilitate investigation.

Forensics Providing the ability to investigate incidents by indexing and searching relevant events.

Reporting Documentation of control sets and other relevant security operations or compliance activities.

ASBs SIEM implementation takes in data from a variety of sources, it can essentially be thought of as a large data warehouse. The problem with any large quantities of data becomes effectively using the data to create intelligence. The data sources currently feeding into SIEM include:

Security Infrastructure Intrusion detection/prevention systems, firewalls, proxies

Network Infrastructure network devices (e.g. Routers and Switches), network services (e.g. DNS, DHCP)

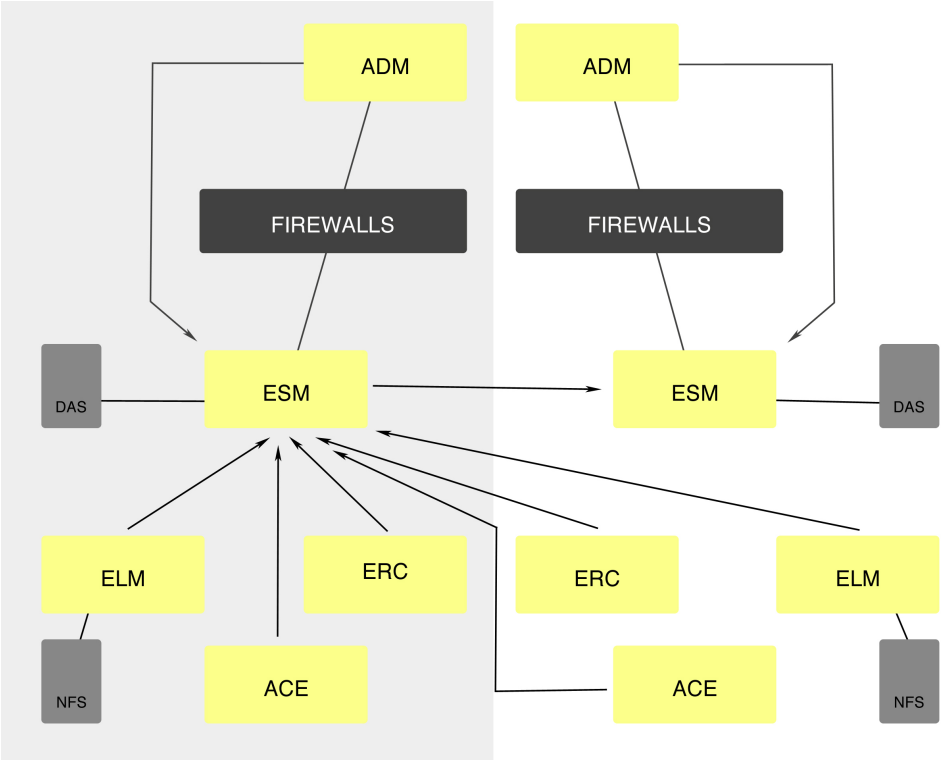

Figure 1: ASB SIEM Architecture

Platforms Operating systems (e.g. UNIX, Windows), DBMS (Database Management Systems)

Applications Over 400 internal applications feed directly into the SIEM environment

Given the sophistication of todays attacks, having all security related data in a centralized location helps the security team identify events of concern faster, analyse the information and incorporate data from different sources more easily and effectively.

4.2 SIEM Architecture

ASBs current SIEM architecture is relatively mature, in terms of set-up and redundancy. In Figure 1 the primary location is on the left, whilst the right hand side represents the secondary location.

4.2.1 ESM Enterprise Security Manager

The ESM is the main component of the SIEM architecture; it is the physical interface that is used to display information and where customized sub-systems can be built in order to perform specific tasks. The physical display is broken down into three distinct portions. The left-hand side encompasses the different incoming data sources. In the centre, is where SIEMs event analysis is displayed, this area is easily customizable in order to fit the needs of the specific task. On the right hand side of the interface, the user can select from a number of filters in order to look for and select distinct types of

data. The filters are extensive, with options available to filter on protocols, traffic and many other characteristics.

The ESM also has a DAS (directly attached storage) attached to it, allowing it to store more records.

4.2.2 ADM Application Data Monitor

The ADM primarily provides Net-Flow data to the ESM. Net-Flow allows us to collect IP Network traffic as it enters or exits the interface. By analysing this type of data in the ESM, we are able to determine the source/destination of the traffic [9].

The ADM is also able to look into the application data and detect sensitive information that is being transmitted inside email attachments, file transfers, HTTP posts etc. You are able to pick up sensitive data types, log them and inform the appropriate parties.

One of the key issues related to the SIEM architecture, is the location of the ADM. As seen in Figure 1 the ADM is placed above all firewalls, this means that all incoming traffic is logged and sent to the ESM, thus creating an enormous amount of data. There are both advantages and disadvantages to this, as this method of collection can give a good indication as to what is going on in the enterprise as a whole but also creates un-needed data.

4.2.3 ERC Event Receiver

The Event Receiver is responsible for collection of events and flow information, from data sources such as those listed in the previous section, including firewalls, applications and work stations. The ERC is able to collect tens of thousands of events per second and store them.

The ERC feeds raw log files directly into the ELM at a set interval. The ESM also retrieves data directly from the ERC at a set interval.

4.2.4 ELM Enterprise Log Manager

The ELM receives the non-indexed and non-aggregated data directly from the ERC. The data here is parsed into a standardized format in order to standardize the data from the various sources it is retrieved from. It also stores the original record so that the activity cannot be repudiated. This enables easy access for compliance monitoring and forensic investigation. Since the original files are not altered, the ELM is able to support chain-of-custody efforts.

The ELM also has a NFS (network file share) attached to it for additional storage needs.

4.2.5 ACE Advanced Correlation Engine

The ACE is fed data directly by the ESM at a set time interval. The ACE is able to monitor close to real time data and aggregate events in order to present them as one in the ESM. The correlation rules have been predefined by McAfee but can also be set and altered in the ESM. ASB has set-up a number of correlation rules in order to customize the SIEM to fit the information security goals.

4.3 Use of Global Threat Intelligence (GTI)

The McAfee GTI framework is a cloud based threat intelligence tool which has the ability to look into multiple types of cyber threats from around the world and allow an organization to make decisions in real time. The data comes in from millions of McAfee products acting in sensors that are deployed across the world in organizations. With every query that is sent, the GTI system is able to learn something new about the subject. This information is combined with other threat vectors such as known malware types to provide a solid world-view of trending threat activity [10]. There are five main types of reputation and categorization services that are implemented as part of GTI. They are:

File Reputation the GTI system is able to look at a file and rate it based on the likelihood that it is malware. This is done through sensors that are deployed across organizations as well as analysis performed by McAfee researchers and cross-vector intelligence tools from email, web and network threat data. The score that is assigned can be used to block or quarantine a file depending on local policy.

Web Reputation The GTI system is able to look at a URL, DNS server or web domain and determine the likelihood of the service being a phishing site, being infected with malware or otherwise malicious. After a score is determined, actions can be taken depending on local policy.

Message Reputation The GTI system is able to determine if an email message is malicious by analysing the message contents many dimensions. It combines this information with spamming patterns and IP behaviour in order to provide a score based on likelihood of the file being malicious. This allows actions to be taken depending on local policy.

Network Connection Reputation McAfee is able to collect billions of IP addresses and network ports, using this information the GTI system calculates a reputation score based on port numbers, protocols and connection requests. The score reflects the possibility of the connection type being malicious. This score is then used in local policy to take action against the network connection.

4.4 Use of Third-party Blacklists

Many organizations maintain and publish blacklists of URL's and IP addresses of known malicious activity. Most of these lists are available online for free to use and implement. The lists generally differ in goals, format and collection methodology. Some lists may target a specific type of malware, whilst others collect a wider variety of data. The lists currently being used and implemented by ASB currently include:

Dyre A watch list for a particular type of malware named Dyre. The malware is used to target users through phishing emails that appear to be from financial institutes [14].

EmergingThreats - an open source platform more than 10 years old for collecting Suricata and SNORT rules. More than 20,000 active users download the rules daily. Provides a watch list based on input from the user community on a wide set of vulnerabilities [6].

Malc0de An open source database of URLs and IP addresses that are hosting malicious executable files.

Palevo A watch list that tracks a worm known as Palevo which transmits using instant messaging, P2P networks and removable drives such as USB media [20].

Zeus Tracker IP addresses and domain names that contain known command and control servers associated with Zeus crime ware. Zeus tracker offers both bad domain name lists and bad IP address lists [3].

5 Current System

5.1 SIEM

The SIEM platform provides a number of pre-built sub-systems in order to help security professionals achieve their goals faster and more effectively. The ASB Information Security team has also built several sub-systems in order to serve a single purpose. In this section I will be looking at several of these sub-systems and analysing the information security goals they help achieve.

The two inbound file sub-systems I will be looking into are primarily created for malware detection, as seen in the Carbanak case study [7], one of the key aspects of the attacks was the use of vulnerabilities in MS Office documents and drive by download attacks. By detecting all the incoming files into the ASB Enterprise, we are able to check if potentially malicious files are being downloaded and/or if they are coming in via the means of email.

5.2 Inbound Exe.

The purpose of the inbound executable subsystem is simply to detect and record all inbound executable files that are coming into the ASB enterprise. This is done by filtering the data type by the unique signature ID that is associated with executable content. The SEM is able to look into the header of the document to detect the file type rather than looking at the extension. This allows detection of files that have been renamed to avoid being captured by low-complexity firewalls.

The general investigation process for the Inbound Exe. Subsystem is to filter once more by "source IP". This allows us to look into the executable files and pick out the files that have potential to be a threat as they originate at a known suspicious or malicious IP address.

From this point forward the processes is largely manual. For data that is coming in via email, the mail content management system must be used to see if the executable did indeed reach the target or if it was blocked at one of the firewalls such as the email security gateway. To support automation of this process, plans are currently in place to integrate mail and web content management logs into the SIEM platform.

5.3 Inbound Office

The inbound Office sub-system is in nature very similar to the inbound executable. The purpose is to record all inbound office documents that are coming into the ASB enterprise. This again is done by filtering for the unique signature ID that is associated with all office documents.

The general investigation process for the Inbound Office documents is very similar to the Inbound Executable. After the user has filtered down for suspicious or malicious IPs, we must look into either

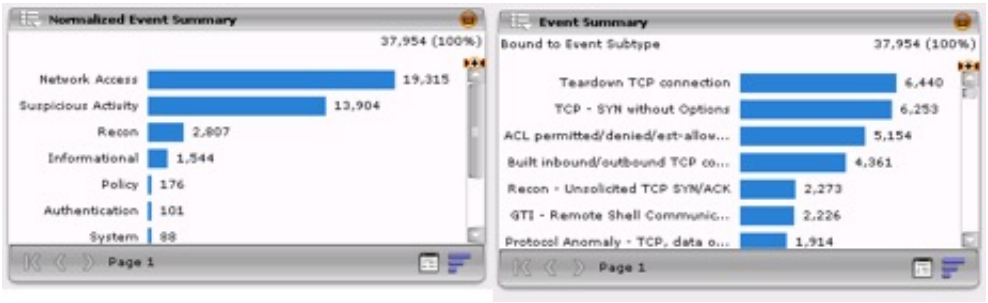

Figure 2: Modules from GTI Systems

the mail content management system or the web proxy/content management system to see if the data was blocked by the firewall before it got to the end user.

The placement of the ASA provides both advantages and disadvantages in data collection. As it is placed above the firewalls, it is able to collect all data that is directed to the ASB Enterprise, this allows a very accurate view of incoming traffic, however as ASB has strict policies in place around office documents, for example, Power-point documents are blocked to a degree from going in or out of the system through Email, however the ASA will pick up this activity regardless and feed it into the SEM. This then creates manual work for the user, as they must investigate further to check at what point the file was blocked or if it reached the end user.

5.4 GTI Inbound and Outbound

The GTI Inbound and GTI Outbound sub-systems are based on the Global Threat Intelligence that is gathered by McAfee. Inbound presents all the data that is coming from a known suspicious or malicious IP and Outbound does the opposite.

The GTI dashboards also normalize to two levels. The first level of normalization provides a high level overview of which category an event belongs as seen in Figure 2. The second level provides a much more detailed explanation as to why the particular event was grouped in this way.

One of the key issues relating to the GTI dashboards is the reasoning behind why an IP address is flagged as suspicious or malicious and the difference between the two. Speaking to experts at ASB revealed that an IP address may potentially be considered suspicious if it is in the same geolocation as another malicious IP. This has the potential to create false positives.

subsectionDefault Summary The default summary sub-system is the first view a user gets once they have logged into the SEM platform. It is preconfigured by McAfee and at the moment does not provide any real intelligence. This is because it simply displays all the data that is available. As seen in Figure 3, the majority of the pie charts displayed do not give real information, the largest portion of each one is listed as “other” meaning it is uncategorised. There is room for drastic improvement

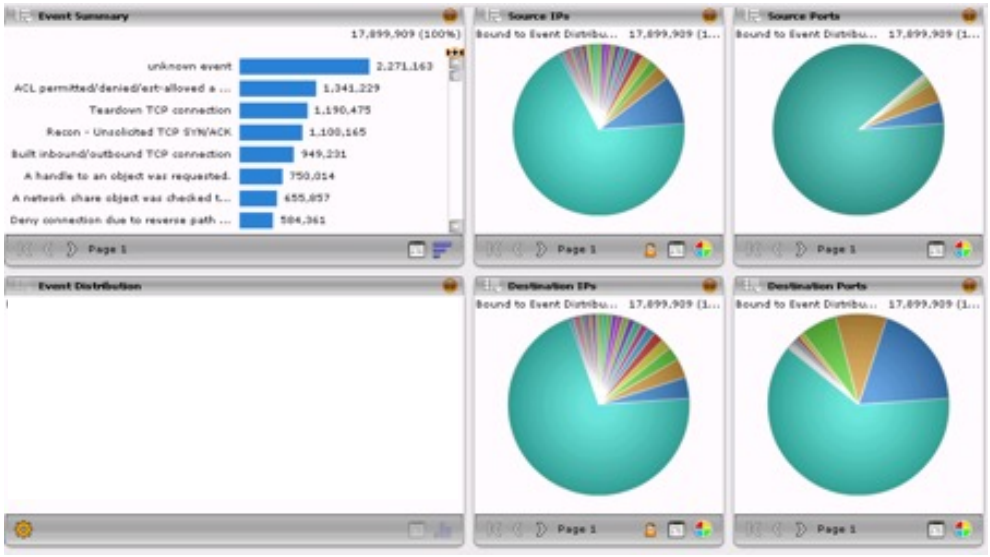

Figure 3: Default Summary System

on this system, so that once a user logs onto the SEM they are able to get a clear idea of what is going on in the organization security-wise.

Figure 4: Outline of Information Security Summary System System

6 Implementation of Threat Overview System

6.1 Recreation of the Default Summary

As previously mentioned, the default summary sub-system is the first view a user gets once they have logged into the SEM platform. My goal for this part of the project was to recreate this system so that it provides useful information and intelligence to the security professionals using it. The sub-system I create is intended to replace the current default summary and provide an overview of what is going on in the ASB enterprise with a single glance. The newly created sub-system will be called, Information Security Summary

In order to do this, I have selected a number of preconfigured dashboards that I believe to be of importance. This selection process was done through conversations with Information Security experts at ASB Bank and asking them, what sub-systems they used the most and which ones were most underutilised. From this information I created the basic outline as displayed in figure 4

As all the information required to investigate a particular event cannot fit onto a single system view, my main sub-system features six modules, each one relating to a different key issue related to information security. There is also be six other sub-systems providing a mode in-depth view into each module coming from the main sub-system, this idea is visualized in figure 4.

The six modules that will be incorporated into the system are:

- Critical File Issues
- Critical Authentication Issues
- Administrator Access Summary
- Severity of Correlated Events
- Global Threat Intelligence Inbound Data

Figure 5: Outline of Information Security Summary System System

- Global Threat Intelligence Outbound Data

Each one of these six modules also has an accompanying sub-system that will allow the user to look into the module in more detail. These sub-systems have initially been taken directly from the current working implementation but in the future have the potential to be expanded and modified as needed.

6.2 Version 1.0 of the Information Security Summary

The core of the Information Security subsystem is displayed in Figure 5, as previously stated it contains six main modules. This subsystem was created with the intention of modifying it and enhancing it as the year goes on. The sub-system may also be modified as business requirements change and/or new threats arise. The six current modules are:

Malicious Incoming Files This module lists potentially malicious files that have been received in the ASB enterprise. It is built from the Malicious File Subsystem that I have created and illustrated in section 6.4

Admin Access Summary This module counts the special privileges assigned to a new login for each of the systems that are listed. By viewing this information a security professional may detect an anomaly if one particular system is showing an excess of assigned privileges. There is vast room for improvement in this module as the data can be cut down to the most critical components so it is easier to view.

Critical Authentication Issues This module lists issues related to authentication. At the moment, it is taken directly from another preconfigured system, but in the future has potential to look at only critical system authentication or a particular system by customizing the filters that it takes into account.

Average Event Severity The module looks at the all of the correlated events over the specified time frame. It then displays the average severity of each correlated event. Currently the severity factors for all correlated events are set to 1. Further research needs to go into this module so it is more representative of ASBs Information Security goals.

Incoming Data from Blacklist IPs This module builds on the preconfigured subsystem of GTI Sourced Threats, it takes the most detailed module from that subsystem and incorporates third party IP watch lists to give a more complete and comprehensive view.

Outgoing Data to Blacklist IPs This module simply does the opposite to that of its neighbour. It builds upon the GTI destined threats system and incorporates the third party watch lists in the destination IP filter, thus allowing you to see all traffic going to a suspected malicious IP. Due to internal ASB research, I have decides to exclude GTI suspicious IPs from the list of incorporated watch lists as it is believed to not provide the required level of confidence.

6.3 Customization of Information Security Summary

Due to the ever-changing landscape of the cyber security world, it is important to keep the information security summary system up-to-date with emerging threats. This vector is already implemented through the use of global threat intelligence and third party IP blacklists but can be taken further by creating custom modules for newly released information related to malicious activity.

An example of such actions would be the creation of a module to combat the Shellshock vulnerability that was discovered in September 2014. IP addresses for those looking to exploit the vulnerability were published and shared online by the Information Security community in efforts to help each other protect themselves from attackers. By incorporating this information into a dynamic watch list that tracks traffic from these incoming IP addresses, we would be able to track and take appropriate action against such attacks.

6.4 Malicious File Subsystem

The Malicious file subsystem is an example of one of the six subsystems that can be built for the main default summary system. For this subsystem, I focussed on the two currently implemented subsystems that were being utilized the most throughout the information security team, the inbound

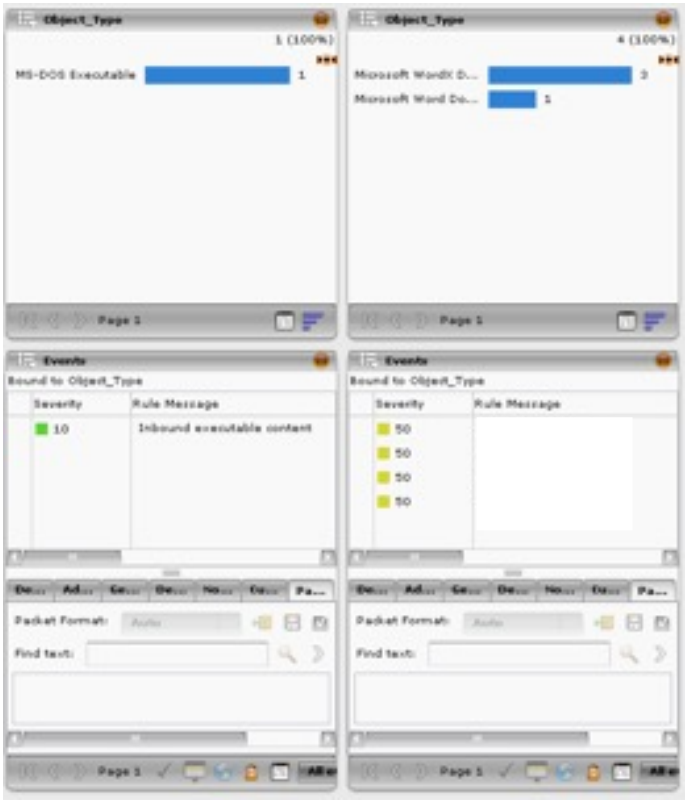

Figure 6: Malicious File System Version 1.0

executables and inbound Microsoft Office documents. These two dashboards are very similar in nature as the primary goal of the two is to look at incoming documents.

The first malicious file sub-system that I created is displayed in Figure 6

This system simply takes the most important data from the two systems in question and displays them in an easy to see manner. The two top modules display the incoming file types of question, whilst the bottom two modules display a breakdown of the events in question.

As mentioned previously, one of the key steps in recognizing a malicious incoming file was filtering the data for malicious source IP addresses. This can be done in the SIEM platform by the use of preconfigured watch lists taken from McAfee Global Threat Intelligence and by the use of third party IP blacklists. As this is an important step in both the systems in question it can also be combined into the new system being created.

This idea was implemented as seen in Figure 7:

Figure 7 displays two modules that are identical to the previous version; they are used to display the incoming executables and office documents. The middle module is used to filter the data through the preconfigured watch lists, such as the ones mentioned previously in this report. By automatically filtering down the data to display the most important aspects, the security professional using the system is able to save time and make more effective decisions in time allocation.

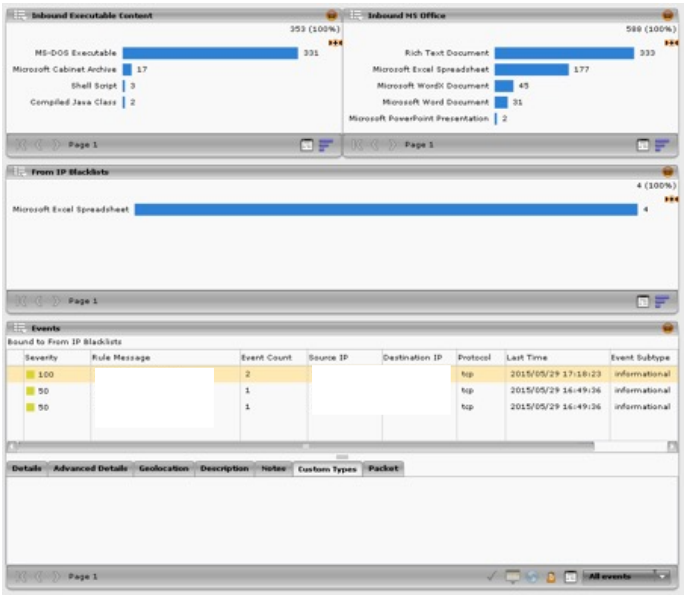

Figure 7: Malicious File System Version 2.0

As seen in 7, in the specified time frame, there were over 300 inbound executable files and over 500 inbound MS Office files, to go through each and every one of these files to determine if they are coming from a malicious source would be both time consuming and extremely tedious. By using this implementation, the events of concern are listed, so that they may be prioritised. The figure shows there were four incoming files from a malicious source. A drill down of these events is visible in the bottom module of the system, the user now simply needs to look at the bottom module and is able to further investigate the event.

6.5 Detection of Communication with C2 Server

As previously stated and seen in the Carbanak case study, once a malicious file has been downloaded on the targets server, it will install malware and try to communicate back with the C2 (Command and Control) server [18]. The lists of malicious IP addresses that are implemented through the use of watch lists in the SEM contain many IP addresses for known C2 servers. In theory, we should be able to look at communications going out to these servers, simply by filtering for destination IP address.

The first step in this process would be to create a list of all the IP addresses that have received content from a malicious IP address, by taking this list and using it as our source IP, we can then filter for a destination IP address which is taken from the preconfigured watch lists. By performing these steps, we should in theory be able to detect if:

1. A user has received a malicious file

2. If the target has communicated back with a malicious IP address.

By doing so, we would be able to see if a machine or particular system in the ASB enterprise has been compromised and escalate the matter accordingly.

However, when trying to implement this procedure, I noticed that the destination IP address that was being displayed was generally not of the host, but that of the Email Security Gateway. This meant that I would not be able to map these to actual machines. The only way to do this would be to first match it to the mail content management system and the web/proxy management system data in order to get the IP address of the targeted host. This in theory should be possible once the mail content management system and web proxy/content management system is integrated into the system, but at the current time can only be done in a manual manner.

7 Literature Survey of Behaviour Profiling Techniques

This section of the report has been done independently from my work at ASB Bank. This was done so that I could get a full understanding of current behaviour profiling methodologies without constraint to the technologies available at ASB Bank. This has given me the opportunity to widen the view I take and therefore effectively evaluate the implementation, in terms of its effectiveness that I put forth to ASB.

7.1 Introduction

The purpose of this section is two-fold. Firstly I am looking at the various behaviour profiling techniques that have been used and implemented in previous years, by doing so I attempt to provide an overview of the ongoing work in this area. Secondly, although the majority of the work related to behaviour profiling has been done in the Intrusion Detection System space, the focus of this paper is not on the implementation of the various systems but rather on the techniques that were used. This is done in an attempt to abstract the techniques used, so that a cyber security professional is able to take certain aspects of the current research and implement their own technologies in an attempt to raise their effectiveness against cyber threats.

This section is organized as follows. Firstly in section 7.2 I outline the multiple data types that may be taken into consideration when designing a behaviour profiling system. In section 7.3, I outline the multiple profile creation methodologies and finally in section 7.3, I discuss the multiple technologies available.

7.2 Possible Datasets/Controls

One of the key components of any behaviour profiling model is the dataset that it is considered. Overall, there are three main data sources possible, user based data i.e. data looking at how a person interacts with a system, system data i.e data looking at how a system responds/behaves and network data, such as NETFLOW data and network layer protocols. Most papers surveyed in this report look at one particular type of dataset. Hybrid systems which can look at multiple types of datasets and aggregate and/or correlate data are available but require features which are not available in most standalone systems.

7.2.1 User Based Data

A number of systems use a user based approach when monitoring for anomalous behaviour. This data represents the many ways a user can interact with a specific system. A key advantage of this

approach is the ability to detect insider misconduct as well as malicious activity coming from inside the organization, such a rogue employee.

Pannell and Ashman [21] proposes a behavioural based IDS focussing on actions that the user takes such as the applications running and keystroke analysis and combining these metrics to detect an intrusion or malicious activity. The specific controls used include:

Applications running By looking at the processes running metric. Also has the ability to determine if a application has been previously run.

Number of windows Measures the number of new processes being created over a certain period.

Websites viewed Measures the number of new sites that are viewed by a user of a period of time.

However the most interesting control looked at in this paper is the use of keystroke analysis through the use of digraph delays, which analyses the time taken between keystrokes. Once a user has typed a certain word multiple times such as CAT. The system is able to determine if the user is legitimate by measuring the time taken between the keystrokes of "C-A" and "A-T". This control however does create the chance of a high number of false positives due the user thinking mid-word or changes in environment such as a new keyboard and overall well-being.

Corney, Mohay and Clark [5] propose a very similar methodology of data collection, however unlike [21] they choose to focus on a single aspect to analyse, the applications used by the user. By doing so they are able to focus on more aspects of this particular activity and look at multiple metrics based around it such as,hour of the day that an application was started, the day of the week an application was used and whether or not the application had been run previously by the same user. A very simple example of this analysis detecting a potential threat would be by looking at a standard working day for an employee of XYZ Corporation. If in general Employee A has checked their email client first thing in the morning every day for the past year and on a particular instance decides to login into a critical system first thing in the morning, this activity could be flagged as anomalous. However a metric such as this would require an extended learning period of only normal behaviour which may difficult in an organization with employees which have constantly changing responsibilities.

Chu et al. [4] propose and test a method for increasing network defence on top of existing solutions in order to identify mis-configurations and breaches of physical security, such as account being compromised. This is done by gathering data from routers and switches across the network. This data is collected from a large-scale ISP network which contains tens-of thousands of routers and switches worldwide. The specific data-types analysed include failed login attempts, user login

access patterns, by looking at an AAA (Authentication, Authorization and Accounting) system data to obtain login IDs, originator IP address and target router IP address.

7.2.2 System Based Data

System data can also be collected in an attempt to monitor for anomalies. Data in this realm generally consists of heat levels, data from HVAC systems and end device data such as CPU cycles and processing power used.

Kim et al. [13] have applied such strategies in order to detect malware that analyses previously unknown energy depletion threats. The framework used consists of a power monitor which collects power levels and builds a database of power consumption history. This type of analysis, which takes into account load on battery levels can be heavily affected by the type of battery being used. After several years of usage, battery levels would not stay relative, therefore raising the false-positive rate.

The work by Abbasi et al. [1] builds on from [13]. They use the system based data approach when looking to monitor anomalies in an embedded system. Due to the nature of safety-critical systems such as military, power plants and industrial automation systems, [1] point out that malware activity can not usually be detected until physical damage occurs. They have chosen to focus on a system that is able to simultaneously able to monitor the total processor power and temperature. By using their methodology they are able to extract the dynamic power, which is shown to be a very good indicator of application activity.

7.2.3 Network Based Data

The network data approach is by far the most common when looking at behaviour profiling techniques. This is largely due to the fact that packet capture can be done easily and provides information rich data that can be extracted and analysed.

Xu et al. [26] look at IP backbone networks in order to determine communication patterns of end-hosts and services. In the study, packet header traces packet-header traces collected on a tier 1 ISP are aggregated into flows based on the source IP, destination IP, source port, destination port and protocol fields. The study is able to show that deep packet inspection as shown by [23] is not needed to identify common communication patterns and extract interesting and/or anomalous behaviour.

McHugh et al. [17] use a network data approach but rather than real time monitoring, choose to focus on forensic analysis by using Net-Flow data, which was originally designed by Cisco Systems. It is supported by a variety of routers and devices for the primary purpose of traffic capture [9]. The information captured within the Net-Flow analysis includes source and destination IP addresses,

protocol ports for TCP and UDP and message type for ICMP. The flow volume is also captured. The reasoning for using Net-flow data was primarily because of the wide usage of the protocol. Privacy concerns over packet inspection were also eliminated by only looking at meta-data/packet headers rather looking into individual packets themselves. One of the downfalls of using Net-Flow data is the stress it puts on components already performing other tasks. To help alleviate this issue, a sampling of data was taken rather than capturing every packet. A better solution to this problem is also proposed, dedicated selection hardware. This dedicated hardware solution is further expanded by Anderson and Arlitt [2] as they describe a method of capturing all packets at 10Gb/s using dedicated hardware. This illustrates the resources required in order to be able to perform full packet analysis.

Additionally, the work done by Chu et al. [4] also encompasses an element of network data as source and destination IP address are observed when looking at the AAA system.

7.3 Profile Creation Methodologies

In order to be able to distinguish between normal and abnormal behaviour, any behavioural profiling methodology must first use a learning phase to create an initial profile. This initial profile provides a basis of understanding of the data, we are then able to compare this profile with additional data in order to determine if anomalous activity is occurring. There are three main ways this can be done.

7.3.1 Rule Creation

The simplest way in which an initial profile may be created is simply by creating rules based on pre-existing research done in the field of information security. Ingham and Forrest [11] showed how this has been applied for a number of years in the field of firewalls by applying a simple block or pass firewall policy. This same methodology has been widely used in the implementation of access control lists [25]

A more complex rule creation methodology can be seen in many commercial SIEM(Security Information and Event Management) systems. Nicolett and Kavanagh [19] survey multiple SIEM technologies and outline how many of them are able to correlate multiple events from across different logs in order to detect suspicious behaviour. Rules can then be applied to these correlations to trigger alerts or take actions when one or more of the correlation rules is satisfied.

7.3.2 Static Profile Creation

The main basis behind the usage of a static profile is that the learning period for a profiling system is performed once. This learning period must encompass normal behaviour and be free from anomalies. This period is then used to decide if current data is normal or anomalous by comparing to the initial

profile. Yeung and Ding [27] have shown that static profiles can be useful when the normal system behaviour does not experience large amounts of variation. The static profiling methodology however is very likely to break down after a certain time-frame as the normal behaviour will have changed over this period. This is also shown by Poo et al. [24]. The static profile methodology was also tested by Corney et al. [5], they carried out a learning phase over a one week period and kept the profile the same over multiple weeks of testing. They were able to conclude that the static profile methodology does not hold up well over any significant period of time as the false positive rate steadily increased the further the testing period was compared to the initial profile period.

7.3.3 Dynamic Profile Creation

Dynamic profile creation is now the norm amongst almost all anomaly based intrusion detection systems. By being able to change the dataset that current events are being compared to, many of the technologies outlined in [22] were able to reduce false-positive rates and increase their anomaly detection capabilities.

Corney et al. [5] outline two different mechanisms for creating a dynamic profile. The first, named the "Growing window" which after the initial week of learning, is extended so that additional events can be added to the profile on a week to week basis. With this methodology, issues that would raise alarms in the static methodology such as software updates can be avoided. An issue that arises from this methodology is the size of the profile, if a user changes roles often the profile may become overloaded with stale history. The second mechanism that is used and tested is the "sliding time window" which involves the width of the profile remaining constant. As new data is added in, old data can be deleted. This is done after training the user profile and generating alerts in the first week. This approach will prevent the user profile from becoming cluttered and filled with historical data.

Pannell and Ashman [21] also use the dynamic profile methodology, however as they were using multiple controls, the nature of the profile was highly dependent on the control being used. Essentially a different profile had to be created and maintained for each control. For applications running, number of windows, application performance and websites viewed the learning period is dynamic as current state data is compared to that of the last hour to the last six hours. For the keystroke analysis the learning has to be done in real time. Each digraph must be compared to the previous in order to determine if there is an intrusion or the user is simply typing a single digraph slower than usual.

7.4 Technologies

The majority of the work done in the are of behaviour profiling is based on IDS solutions and in particular, anomaly based intrusion detection. [22] provides a comprehensive overview of Anomoly-based IDS solutions, Whilst, [5], [21], [24] and [4] all focus on a single IDS solution. SIEM solutions are slowly beginning to now have the capability of implementing behaviour profiling also, [19] studies SIEM solutions in-depth and compares the profiling capabilities of multiple platforms.

8 Behaviour Profiling System Design Decisions

As the literature survey was done independently from the implementation at ASB, in this section of the report I outline the design decisions that resulted directly from the literature survey, as well as other important factors taken into consideration when designing the behaviour profiling system.

8.1 Findings From Literature Survey

8.1.1 Technology

As the project brief indicated the need to grow and mature the SIEM platform, the McAfee SIEM solution was the obvious choice of technology for this system.In the literature survey, several possible technologies were identified, which had the capability of implementing behaviour profiling. These technologies can be categorized into the main groups, including custom scripts, Intrusion detection systems (IDS) and Security Information and Event Management (SIEM). Due to the restraints of the project, only two of the three were considered as viable solutions. These were the use of the custom scripts and the use of the McAfee SIEM solution. Custom scripts have the advantage of allowing a user to accomplish a wider array of tasks as the limitations of any pre-existing software tool can be overcome. However due to the nature of the sensitive data, time restraints and initial brief, it was the McAfee SIEM solution was decided to be the best choice of implementation for this project and ASB going forward.

Correlation Rules From the research conducted into behaviour profiling techniques and the ASB SIEM environment, I was able to determine that the best way to implement behaviour profiling into the current SIEM platform would be by the use of the advanced correlation engine, that was previously discussed in the report. The ACE allows the user to create and manage correlation rules. These rules tie together events from multiple sources to provide a higher level event. One example of this is the rule named "Login - Brute Force Login Attempt from a single source". This rule has the ability to match multiple failed authentication events that originate from the same source IP. By doing so, potential brute force attacks, which are used to try and guess the password of a user can be blocked. Severity levels can be set-up with these rules so that events which pose a greater threat to the environment are more visible than those that are not.

8.1.2 Data Used

User based data was determined to be the most important when considering a behaviour profiling approach. By looking at the user dataset, we are able to determine the normal day to day behaviour

for a particular user and user group. Network data was also looked at to determine the method of connectivity to the ASB system and to determine the ingoing and outgoing network behaviour was not malicious. For the purposes of profiling a users behaviour, system data was determined to be not useful. As seen in the survey, system data is generally analysed in embedded systems and data originating from HVAC or similar systems. I felt this datatype would be useful for the purposes of this project.

8.1.3 Profile Creation Methodology

The profile creation methodology that is used in this system is compromised of a rule-based profile and dynamic based profile. From the survey, it can be seen that a static profile is of minimal use over a long period of time and therefore it was not considered to be a viable option. In an ideal situation, by having a rule-based methodology we are able to balance some of the negative aspects of a dynamic approach. An example of this is an attacker that works slowly, making incremental changes to the user accounts behaviour so that it is determined normal by the learning profile, by including a rule-base if an abnormal action is ever taken, regardless of previous behaviour it will be flagged.

8.2 User Group Definition

A key phase of this project was the selection of a user group to conduct behaviour profiling. Eventually this method of profiling will be carried out throughout the ASB Enterprise. However for the purposes of this project, the goal was to select one key user-group and make that group the focus. Several user groups were considered and these including:

Executive Leadership Team (ELT) THE ELT consists of the CEO and those who report directly to the CEO. These accounts contain vast amounts of confidential information that could potentially damage the company and/or help competitors if leaked. The ELT is also very mobile, therefore tracking where and when the accounts are being accessed from is critical and could be an early indicator of a compromise. The personal assistants (PA's) are also included in this group as many of the resources for the ELT are also on the accounts of the PA's, if a PA account was to be compromised it would do the same damage, if not more than an ELT account.

Information Security Team The IS team has access to a wide range of applications and data sources within the ASB Enterprise, as they are tasked with protecting the systems, if an IS account was to be compromised, there would be potential to use the information gained to conduct a targeted attack against the enterprise

System Administrators The SysAdmin group is in charge of ensuring that the system is available and functioning as it should. The roles of the SysAdmins vary significantly both in terms of tasks and privelage levels. A SysAdmin account has the potential to shut down particular systems. If such an account was to be compromised, large amounts of confidential data could be lost and potentially the infrastructure of the ASB Enterprise could be seriously affected.

Taking these considerations into account, the decision was made to select the ELT as the primary user group for this project. The system administrator group was simply too large to focus on and the tasks they carried out were too varied. This would make it very difficult to decide what constituted "normal" behaviour. In terms of the IS Security team, it was simply deemed that profiling of the ELT would be more beneficial, this is because the IS accounts are generally monitored in other means, where as there is no dedicated monitoring portal for the ELT.

8.3 Usability Heuristics

There were considerations that needed to be taken into account when considering the design of the behaviour profiling system. These considerations include:

Target User Group When designing the system, it was important to consider the end-user of the system and their knowledge of the ELT and Information Security in general. The end-users of this particular system are the members of the ASB Information Security team. This allowed me to design the system with the premise in-mind that all those who use it would be highly skilled and know signs of potential abnormal activity and misuse. As many members of the team have been at ASB for a reasonable amount of time, it also meant that the end-users would know the subjects and their general day-to-day behaviour quite well. From these two observations, I was able to conclude that if the system displayed the daily behaviour of a particular user in the ELT, the security professional looking at it, would be able to deduce if the behaviour was normal or not.

Time to Load Due to the limited resources that are available, it is important to ensure that the system has a reasonable loading time. By conducting internal research at the bank and taking the tools limitations into account it was decided that the ideal loading speed for the system should be 10seconds with an upper-limit of 30seconds. IF any modules, on average were to take longer than the upper-limit they would be proven to be ineffective. Although an upper-limit of 30seconds is relatively low in such a system, I was able to find that if the initial module was to take longer than this limit, any analysis on that module would have an approximate loading time of greater than one minute. This meant that if a security professional wanted to

explore several aspects in detail, the loading time would simply be too slow to be effective.

Future Proof As the SIEM solution continues to evolve and mature, it is important to design a system that will automatically become more effective as the system matures. The best example of such a trait is with the correlation rules that have been implemented. Many of the default rules have been restructured and replaced by ASB specific rules. As this refinement continues and more data-sources are fed into the system, it was important that the proposed solution make full use of the added refinement with minimal human intervention

9 Implementation of the Behaviour Profiling System

Following discussion with the other members of the Information Security team and additional internal research, the decision was made to create two separate subsystems, similar to the style of the Information Security Summary System. This was done as several of the modules I was trying to implement simply did not work well with user-group larger than 1. Therefore two modules were created to overcome this obstacle.

9.1 Profiling of ELT

The first system as seen in Figure 8 is intended to profile the behaviour of the ELT and PA group. The intention of this system was to utilize the power of the SIEM platform and in particular the ACE. There are three main sections in this particular system. They are:

Correlation Rules Triggered This section of the system looks solely at the ACE to detect all correlation events that have been triggered by the user-group. By looking at all of the correlation events, rather than a subset of, we are able to future-proof the system. This will mean that as new correlation events are added, they will continue to automatically appear in this system if triggered

Correlation Rules Hit Stacked by User This module tracks the correlation rules triggered and displays them in a bar graph over the period of time that the system has been initialised for. By looking at the trends of the correlation rules and stacking by user-ID, we are able to see which users are potentially being targetted and focus on these users to ensure that there is not a present threat.

Event Source Users This module lists the event count taken directly from the "Correlation Rules Hit stacked by User" module. By tracking the amount of correlation events triggered by a single user, we are able to see trends and once again, see which users are most active.

Event Source IP This module allows us to see, the originating IP address of all the correlation rules triggered. If an external IP address appears on this list, it is likely to be a greater threat,

Events This module lists all the correlation events that were triggered for the ELT user group. By tracking the rules triggered we are able to get an understanding of potential attacks as they come up. For example if several rules concerning brute force attacks from an overseas location appear, we are then able to provide appropriate countermeasure.

Average Severity Ratings These modules provide average severity ratings for several aspects. As events are generated, a severity is assigned to each based on pre-existing knowledge, by tracking this severity and comparing to historical data, we are able to get a greater understanding of the current situation

Average Event Summary for Corr. Rules Looks at the average severity rating for each correlation rule that is triggered. Highlighting the rules that are most severe and require immediate attention.

Average Source IP Severity Looks at each source IP and determines the average severity of all the events occuring from it. If a severity rating for a single internal IP is high, it may need further investigation. If an event severity from an external IP is high, it may need further investigation or alteration of firewall rules may be needed to block this IP

Average Source User Severity Looks at the events being generated by each source user and displays the average severity rating. Again, if a severity rating is relatively high, further investigation into the account will be needed to ensure that there is no compromise.

Application and Remote Access Looks at the applications accessed and the remote access behaviour of the user group

Applications The ETL has a specific set of applications that are likely to be accessed. If a suspicious application appears in this list, that is generally unused by the user group, further investigation may be needed.

Remote Login Events Remote Access events can be one of the earliest sign of compromise. This module is able to look at where the user-group is accessing the ASB system. As the ELT's schedules are available to the Information Security team, if a remote access occurs from a suspicious location, further investigation may be needed.

9.2 Singular Account Profiling

The second of these two systems is treated as an additional source of information to the first. This system has the ability to analyse any one user account across the ASB Enterprise. If a user account is suspected of being compromised or an employee is under suspicion of misusing their account, this system has the capability to track the users behaviour to conduct an investigation. The modules in this system can be categorized into seven categories. They are:

Trends Looks at the activity by the user over the specified time period. By checking the activity of a user, we are able to see the time that the user first accessed the system, times of heavy

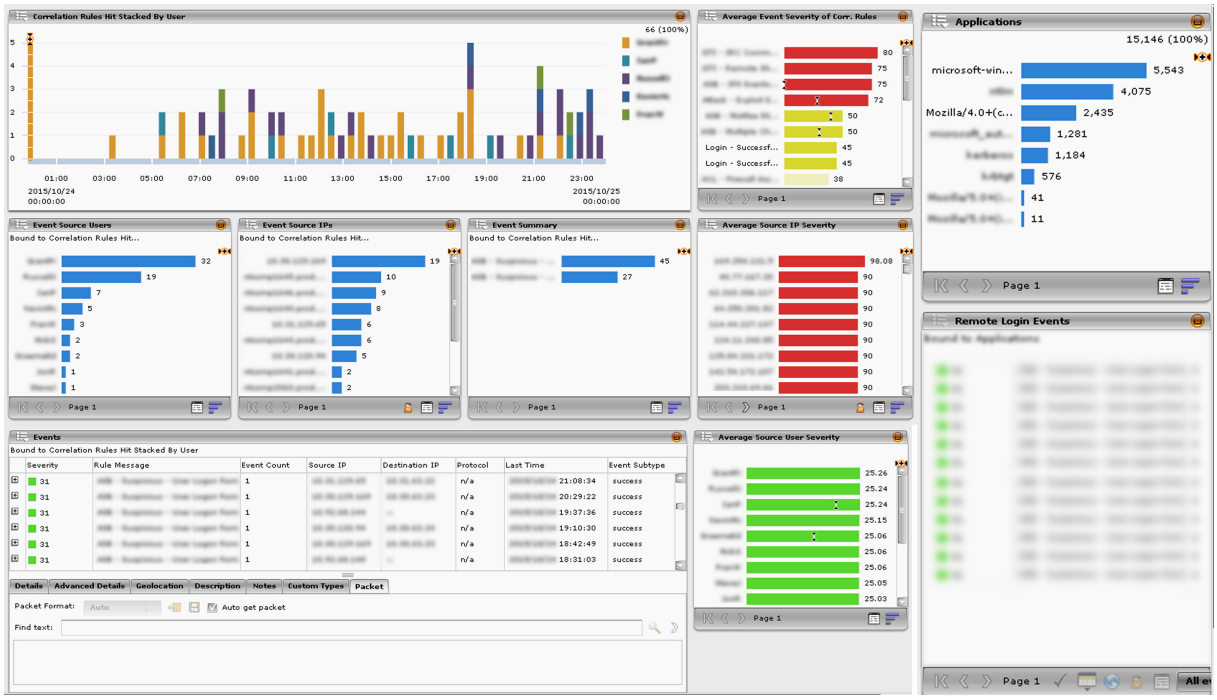

Figure 8: Behaviour Profiling System based on the ELT

use and the time the user has exited the system.

Login Events This module looks at the successful login events for a use throughout the ASB system. This event type was the best way to model the day to day activity of a particular user. By comparing the event curve with historical data, we are able to see if there are any unusual trends for a particular day.

Login IP's The login IP's module simply lists all the IP addresses which the particular user has logged in from. As this system is generally set to output data of a single day, multiple IP address entries could cause a red flag. Particularly if these are a mixture of internal and external addresses.

Applications The applications module lists all the applications that a user has accessed over the specified time. It is the same in nature as the application module in the ELT system

Correlation Events The correlation events module lists all the correlation rules that a user has triggered. This is an excellent way to detect potentially suspicious behaviour, depending on the rules triggered and the amount of rules triggered.

Communication The communication module looks at the Event Network Graph. This module has the ability to look at the source IP addresses and match them to the destination IP addressed or vice-versa by dragging one of the IP address circles to the centre of the diagram. By looking

at the communication for a particular user, we are able to track potentially suspicious events. This module could potentially also be used for detecting communication with a C2 server, as mentioned previously in the report.

All Events As the user-group of this system is singular, it makes sense to list all events that a user triggers. By doing so we may be able to spot a potentially malicious event that has not been picked up by the other modules. This modules success rate however is highly dependant on the amount of events generated. If a user is very active, it will be very difficult to analyse through this module.

Remote Access Events The remote access module is the same in nature to the module in the ELT system. Remote access can be an early sign of a compromise, particularly when looking at a single user. By using other resources that are available to the Information Security team, we are able to see if a user is at the office, working from home or working from overseas. If the remote access patterns do not match up, further investigation is warranted.

Dials The dials reppresent very quick facets of detail that can tell us about the activity of the user.

Total Events Measures the total number of events that are generated by a user over the specified time period.

Average Severity Count Measures the average severity of the user, based on the events being generated.

Average Risk Measures the risk of the user, based on the events being generated.

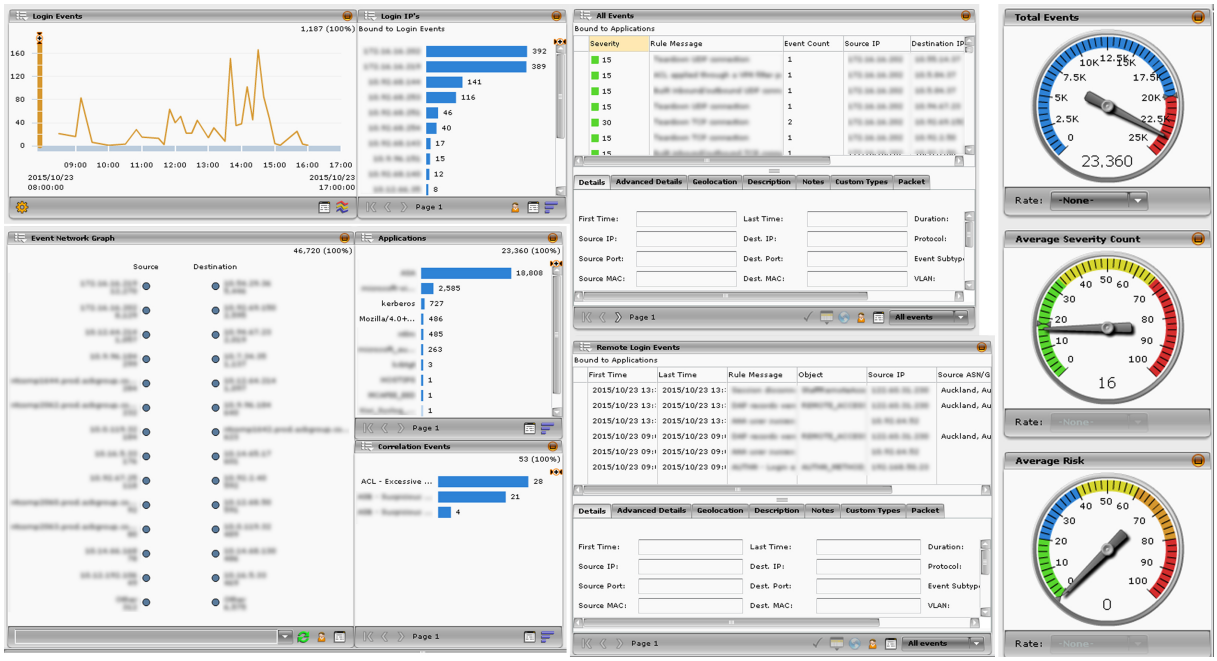

Figure 9: Behaviour Profiling System based on a Single User

10 Refinement and Feedback Loop

Over the course of the project, apart from the presentations at the University of Auckland, I have also presented my work to the ASB Information Security team. Through the course of the year three main meetings have taken place and a number of smaller meetings with one or two members of the IS team have also taken place.

The purpose of these presentations has been to present the work I have conducted at the bank. As all the systems that I have been created are intended for use by the IS team, it allowed me a chance to show the team the functionality built into the systems so that they may put them to use. The presentations also provided a chance for the IS team to provide me feedback so that additional improvements to the system could be made.

One of the key findings from these meetings was the decision to select the Executive Leadership Team (ETL) as the prime user group for this project.

10.1 Use of the Implemented Systems

Both systems that have been implemented at ASB Bank are now in use by the IS team. The first system, the Information Security summary, along with its accompanying subsystem has been in use for a total of five months to date and has in retrospect replaced the Default Summary page that it was intended to replace. The system is primarily used as a gateway for further investigation as it is able to display the most critical events that must be attended to. The malicious file module has

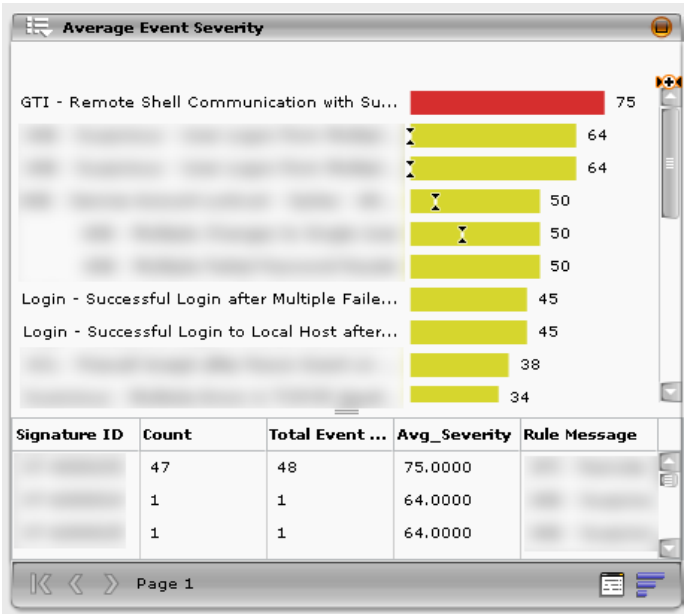

Figure 10: A module of the Information Security Summary v2.0

been the most successful as it can provide early signs of a potential malware attack.

10.1.1 Redesign of the Information Security Summary

Due to feedback from the Information Security team, two key changes were made to the Information Security Summary System. These include:

Inclusion of Events The first change is for a purely usability standpoint. By including the source events in each particular module, the end user is able to quickly see additional details from each particular module. This reduces the number of clicks needed to access data and makes more data available in a single glance.

Comparison to Historical Data By comparing the current days data to historical data, we are able to see the change that has occured on any given day. This becomes a useful peice of information as we are able to clearly see deviations from normal behaviour and adds value to the system.

Both of these changes can be seen in Figure 10, which represents a single module of the new Information Security Summary System.

11 Key Issues and Challenges

11.1 Familiarity with and Size of Dataset

The Current SIEM implementation consists of terabytes of data. There is well over a million events that are recorded each day. In order to gather intelligence from the dataset, it is important to understand the underlying logic of the SIEM. It is important to understand the data sources as well the different fields that are presented. This is a crucial step in determining if the data is of use or is simply background noise that can be ignored.

Due to the size of the dataset, performing analysis on historical was very difficult and in some cases impossible. Running queries over a period of over a month and in some cases a week required a waiting period anywhere between five minutes to an uncalculated amount of time. This meant the subsystems I created could at most only look at data for the current or previous day.

11.2 Missing Dataset

As mentioned previously in my report, the process of adding data into the SIEM platform can at times be very complicated and time consuming if the data is source is not directly supported. The parsing of this data to make it compatible for SIEM is an ongoing process at ASB Bank and is prioritized from most important to least. Currently data from the mail content management system and web proxy/content management system are not implemented in SIEM. When implementing the automatic detection of communication with a C2 server as outlined in 7.3.1 this was a very apparent issue. In order to truly get the full picture of what is happening security wise in the ASB enterprise it is essential to have this data feeding into the SIEM platform and is a prioritised task for the Operational Security team.

11.3 Access Privileges

Due to the confidential nature of the dataset, my profile for the ASB SIEM platform did not allow full access to all the tools and utilities available. This was done as a security measure by the bank and is common practice throughout corporations. This is a measure that has to be taken in order to be compliant but posed several challenges as I completed this project. PArticularly, as I was a part of the Information Security and the task of SIEM integration was assigned to the Operation Security team, there were certain aspects of the platform that I was unable to explore.

11.4 Limitations of Tool

As the primary objective of this project was to increase the intelligence output of the SIEM tool, I was limited to only using this tool. For some aspects of the project, it is possible that custom scripts would have been slightly more efficient. Due to the loading time of certain modules, these modules could not be implemented. Furthermore, there were ongoing database issues that restricted the speed of the platform and these had to be taken into consideration when designing the system.

12 Conclusions

In this report I have outlined my process for enhancing the intelligence capabilities of the ASB SIEM solution. The first system I have created, the Information Security Summary is now in use and has effectively replaced its predecessor the Default Summary System. The second system I have designed has been based on extensive research in the field of behaviour profiling. Using this research I have come up with and recommended a behaviour profiling methodology which has been implemented at ASB Bank.

In terms of future work, I would like to return to this project and test the effectiveness of the solutions that I have implemented. As the second system, encompassing behaviour profiling was only recently put in production, I was unable to receive criticisms and recommended changes. Furthermore, enhancement of correlation rules would likely have a positive impact on the results gained from the behaviour profiling system. The current approach relies on static thresholds, but the SIEM platform has the capability to implement statistical thresholds, however I was not able to test these due to privilege levels.

From the feedback received thus far from the end users of the solution, the Information Security team at ASB bank, the systems have been able to save a significant amount of time by highlighting the most important factors of concern. An official handover of this project will occur, from that time forward, members of the Information Security team will continue to upgrade and change the systems they see fit and in accordance to the cyber security landscape.

References

[1] Zeinab Abbasi, Mehdi Kargahi, and Morteza Mohaqeqi. Anomaly detection in embedded systems using simultaneous power and temperature monitoring. In *Information Security and Cryptology (ISCISC), 2014 11th International ISC Conference on*, pages 115–119. IEEE, 2014.

[2] E Anderson and M Arlitt. Full packet capture and offline analysis on 1 and 10 gb networks. 2006.

[3] Dennis Andriesse and Herbert Bos. An analysis of the zeus peer-to-peer protocol. Technical report, Technical report, VU University Amsterdam, 2013.

[4] Jie Chu, Zihui Ge, Richard Huber, Ping Ji, Jennifer Yates, and Yung-Chao Yu. Alert-id: analyze logs of the network element in real time for intrusion detection. In *Research in Attacks, Intrusions, and Defenses*, pages 294–313. Springer, 2012.

[5] Malcolm Corney, George Mohay, and Andrew Clark. Detection of anomalies from user profiles generated from system logs. In *Proceedings of the Ninth Australasian Information Security Conference-Volume 116*, pages 23–32. Australian Computer Society, Inc., 2011.

[6] LLC Emerging Threats Pro. Emerging threats open source overview, 2015.

[7] Kaspersky Lab HQ. Carbanak apt: The great bank robbery. Technical report, Kaspersky Lab HQ, February 2015.

[8] Cloppert M. J. Hutchins, E. M. and R. M Amin. ntelligence-driven computer network defense informed by analysis of adversary campaigns and intrusion kill chains. Technical report, Lockheed Martin Corporation.

[9] Cisco Inc. Introduction to cisco ios netflow, 2012.

[10] McAfee Inc. Mcafee global threat intelligence. Technical report, IBM, 2010.

[11] Kenneth Ingham and Stephanie Forrest. A history and survey of network firewalls. *University of New Mexico, Tech. Rep*, 2002.

[12] Ponemon Institute. 2014 global report on the cost of cyber security. Technical report, HP Enterprise Security, 2014.

[13] Hahnsang Kim, Kang G Shin, and Padmanabhan Pillai. Modelz: monitoring, detection, and analysis of energy-greedy anomalies in mobile handsets. *Mobile Computing, IEEE Transactions on*, 10(7):968–981, 2011.

[14] Mueller L. Kuhn, J. and L Kissem. The dyrewolf: Attack on corporate banking accounts. Technical report, IBM, April 2015.

[15] ASB Bank Limited. Our history, 2015.

[16] Intel Security McAfee. Net losses: Estimating the global cost of cybercrime, economic impact of cybercrime 2. Technical report, McAfee, Intel Security, June 2014.

[17] John McHugh, Ron McLeod, and Vagishwari Nagaonkar. Passive network forensics: behavioural classification of network hosts based on connection patterns. *ACM SIGOPS Operating systems review*, 42(3):99–111, 2008.

[18] Ghita Mezzour, Kathleen M Carley, and L Richard Carley. An empirical study of global malware encounters. In *Proceedings of the 2015 Symposium and Bootcamp on the Science of Security*, page 8. ACM, 2015.

[19] Mark Nicolett and Kelly M Kavanagh. Magic quadrant for security information and event management. *Gartner RAS Core Reasearch Note (May 2009)*, 2011.

[20] palevotracker.abuse.ch. Palevo tracker: Home, 2015.

[21] Grant Pannell and Helen Ashman. Anomaly detection over user profiles for intrusion detection. 2010.

[22] Animesh Patcha and Jung-Min Park. An overview of anomaly detection techniques: Existing solutions and latest technological trends. *Computer networks*, 51(12):3448–3470, 2007.

[23] Jignesh D Patel. Algorithms for deep packet inspection. 2012.

[24] Danny Poo, Brian Chng, and Jie-Mein Goh. A hybrid approach for user profiling. In *System Sciences, 2003. Proceedings of the 36th Annual Hawaii International Conference on*, pages 9–pp. IEEE, 2003.

[25] Henk CA Van Tilborg and Sushil Jajodia. *Encyclopedia of cryptography and security*. Springer Science & Business Media, 2011.

[26] Kuai Xu, Zhi-Li Zhang, and Supratik Bhattacharyya. Internet traffic behavior profiling for network security monitoring. *Networking, IEEE/ACM Transactions on*, 16(6):1241–1252, 2008.

[27] Dit-Yan Yeung and Yuxin Ding. Host-based intrusion detection using dynamic and static behavioral models. *Pattern recognition*, 36(1):229–243, 2003.