INTRODUCTION

1.0 Overview:

As the threat posed by massive cyberattacks (e.g., ransomware and distributed denial of service attacks (DDoS)) and cybercrimes has grown, individuals, organizations, and governments have struggled to find ways to defend against them. In 2017, ransomware known as WannaCry was responsible for nearly 45,000 attacks in almost 100 countries. The explosive impact of cybercrime has put governments under pressure to increase their cybersecurity budgets. United States President Barack Obama proposed spending over \$19 billion on cybersecurity as part of his fiscal year 2017 budget, an increase of more than 35% since 2016.

The cybercrime underground has thus emerged as a new type of organization that both operates black markets and enables cybercrime conspiracies to flourish.In contrast, cybercrime networks are lateral, diffuse, fluid, and evolving. the threat posed by the rise of highly professional network-based cybercrime business models, such as Crimeware-as-a-Service (CaaS), remains mostly invisible to governments, organizations, and individuals.

1.1 Motivation:

To achieve this goal, we propose (1) a data analysis framework for analyzing the cybercrime underground,(2) CaaS and crime ware definitions, and (3) an associated classification model. In addition, we (4) develop an example application to demonstrate how the proposed framework and classification model could be implemented in practice.

We then use this application to investigate the cybercrime underground economy by analyzing a large dataset obtained from the online hacking community. By taking a design science research approach, this study contributes to the design artifacts, foundations, and methodologies in this area. Moreover, it provides useful practical insights to practitioners by suggesting guidelines as to how governments and organizations in all industries can prepare for attacks by the cybercrime underground.

1.2 Objectives:

This project collected data from the largest hacking community and did not consider other hacking communities. Future studies will therefore need to generalize our findings by investigating a wider range of hacking communities. Second, this study has focused on the CaaS and crimeware available in the cybercrime underground, but much in-depth analysis remains to be done on the configurations of cybercrime networks.

1.2.1 Advantages:

- Compelling and relevant content will grab the attention of potential customers and increase brand visibility.
- You can respond instantly to industry developments and be seen as 'thought leader' or expert in your field.
- This can improve how your business is seen by your audience.
- Positive feedback is public and can be persuasive to other potential customers.
- Negative feedback highlights areas where you can improve

The goal of the proposed framework is to "investigate the cybercrime underground economy." This Crimeware marketplaces have three key elements, namely

1.actors (e.g., coders, operators, or buyers),

2. value chains, and

3.modes of operation (e.g., CaaS, pay-per-install, crimeware toolkits, brokerage, or supplying data). Data mining is a technique used to mine out patterns of useful data from large data sets

1.2.2 Disadvantage:

- ➤ It is not secured process.
- Over under traction is invents
- Download files, time is invited

The cyber-crime differs from general crime in many ways; we need to conduct a variety of analyses using a large data set. A previous study proposed a data mining framework for crime, dividing crimes harmful to the general public into eight categories:

A DATA ANALYTICS APPROACH TO THE CYBERCRIME UNDERGROUND ECONOMY

1.Traffic violations. 5.Gang/Drug Offenses.

2. Sex Crime. 6. Arson.

3. Theft. 7. Violent crime.

4. Fraud. 8.Cyber-crime.

Although this previous study explained how data mining techniques could be applied to crime analysis, it did not consider the specific features of cyber-crime.

1.3 FEASIBILITY STUDY:

The feasibility of the project is analyzed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirements for the system is essential.

Three key considerations involved in the feasibility analysis are,

- ➤ ECONOMICAL FEASIBILITY
- TECHNICAL FEASIBILITY
- SOCIAL FEASIBILITY

1.3.1 ECONOMICAL FEASIBILITY

This study is carried out to check the economic impact that the system will haveon the organization. The amount of fund that the company can pour into the research and development of the system is limited. The expenditures must be justified. Thus the developed system as well within the budget and this was achieved because most of the technologies used are freely available. Only the customized products had to be purchased.

1.3.2 TECHNICAL FEASIBILITY

This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. This will lead to high demands on the available technical resources. This will lead to high demands being placed on the client. The developed system must have a modest requirement, as only minimal or null changes are required.

1.3.3 SOCIAL FEASIBILITY

The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity. The level of acceptance by the users solely depends on the methods that are employed to educate the user about the system and to make him familiar with it.

LITERATURE SERVEY

Cybercrime has undergone a revolutionary change, going from being product-oriented to service-oriented because the fact it operates in the virtual world, with different spatial and temporal constraints, differentiates it from other crime taking place in the physical world As part of this change, the cybercrime underground has emerged as a secret cybercrime marketplace because emerging technological changes have provided organized cybercriminal groups with unprecedented opportunities for exploitation The cybercrime underground has a highly professional business model that supports its own underground economy. This business model, known as CaaS, is "a business model used in the underground market where illegal services are provided to help underground buyers conduct cybercrimes, such as attacks, infections, and money laundering in an automated manner,". Thus, CaaS is referred to as a do-it-for-me service, unlike crimeware which is a do-it-yourself product. Because CaaS is designed for novices, its customers do not need to run a hacking server or have high-level hacking skills. Consequently, the CaaS business model can involve the following roles: writing a hacking program, performing an attack, commissioning an attack, providing an attack server (infrastructure), and laundering the proceeds. Sood and Enbody have suggested that crimeware marketplaces have three key elements, namely actors (e.g., coders, operators, or buyers), value chains, and modes of operation (e.g., CaaS, pay-per-install, crimeware toolkits, brokerage, or supplying data). Periodic monitoring and analysis of the content of cybercrime marketplaces could help predict future cyber threats. In criminology, routine activity theory (RAT) is used to explain the causes of crime, both general criminal activity and cybercrime.

According to this theory, three elements are necessary for crimes to be committed:

- (1) a likely offender,
- (2) a suitable target, and
- (3) the absence of capable guardians against crime.

In a cybercrime context, the "likely offenders" are motivated sellers and potential buyers in the underground market, and the "suitable targets" are the targeted vulnerable organizations. The "absence of capable guardians against crime" is due to organizations failing to take preventive measures against cybercrime. Two types of product or service are available in the cybercrime underground. The first can be either CaaS or crimeware

that are related to attack strategy, for example, phishing, brute force, or DDoS attacks, or can be used for spamming or creating botnets, exploits, ransomware, rootkits, or Trojans. Attack strategies often exploit system vulnerabilities such as application loopholes. In addition, social engineering attacks exploit human vulnerabilities. The most well-known example of such an attack is the use of a "secret question" for password recovery: attackers check into the user's background to guess the secret question and hence steal the account. Examples include encryption and virtual private network (VPN) services, crypters, and proxies. From the perspective of RAT, the likely offenders are attackers motivated to attack organizations or products that constitute a suitable target. If such targets are attacked, however, both the targets and those who supply their cybersecurity products become aware of the vulnerabilities that made the attack possible, leading them to apply security updates to their software.

These updates can be seen as capable guardians against crime, and the preventive measures taken can be identified by looking through each program's version history. However, this is not the end of the matter, because the attackers will then develop and sell new versions of their hacking tools to combat the guardians, thus re-establishing the third RAT condition, the absence of capable guardians against crime. Such events can also be identified by the version numbers of the hacking tools sold on the black market: since it is an online marketplace, attackers must give detailed explanations to retain their customers' confidence.

2.1 Comparative analysis:

For a number of years, policy-makers at the highest levels have been expressing their concerns that insecure information systems threaten economic growth and national security. As a result of theseconcerns, a complex and overlapping web of national, regional, and multilateral initiatives has emerged. A recent publication offers a compilation and analysis of cybersecurity efforts in fourteen countries.76 The International CIIP Handbook provides an overview of issues of high importance in the field of critical information infrastructure protection (CIIP), serves as a reference work for the interested community, and provides a basis for further research by compiling relevant material. In this chapter, the main findings of this volume are presented. We are focusing mainly on five focal points of high importance that emerged from a cross-comparison of country surveys:

1. Critical Sectors: This section compares critical infrastructure sectors as identified by

therespective countries. To look at the concept of critical infrastructures is important, becausecybersecurity has emerged in parallel to CIP in all countries.

- **2. Organizational Overview**: The second part of this chapter provides an overview of important public actors in the national cybersecurity organizational framework. It only characterizes the specific responsibilities or public actors at the state (federal) level (such as ministries, national offices, agencies, coordination groups, etc.). Public actors at the lower state level and private actors (companies, industry, etc.) are not taken into account.
- **3. Early Warning Approaches**: The third section describes national organizations responsible for cyber-early warning, namely cybersecurity-related information-sharing organizations such asCERTs (Computer Emergency Response Teams), ISACs (Information Sharing and Analysis Centers), etc.
- **4. Current Topics in Law and Legislation:** The development of effective regulations, laws, and criminal justice mechanisms is essential in deterring virtual abuse and other offences against the information infrastructure. Moreover, a strict regulation may create trust in the new ICT and encourage the private sector and individuals to make better use of e-Commerce or e-Government services.
- **5. Research and Development**: The last section gives an overview of recent efforts in Research and Development (R&D) concerning cybersecurity.

2.2 RESULTS AND INFERENCE DRAWN:

The data analysis step of the proposed framework involves four steps. Here, we report the data analysis results: CaaS and crimeware classification and market trends, cybercrime market dynamics, and potential hacking targets.

CaaS and Crimeware Classification and Market Trends

Here, we evaluate the accuracy of the proposed classifications. Specifically, we analyze the CaaS and crimeware trends between 2008 and October 2017 based on these classifications, the most common classes overall were botnets (17%) and exploits (17%). The most popular classes in 2017 were botnets (33%), VPN services (20%), exploits (13%), and brute force attack services (7%).

To validate our classification model, we used a confusion matrix, a common method of calculating classifier output accuracy. The training and testing datasets comprised 300 and 700 items, respectively. This gave an accuracy of 82.6% with a 95% confidence interval of (70.74%, 81.24%) for identifying the risks posed by CaaS- and crimeware related messages.

EXISTING AND PROPOSED SYSTEM

3.1 Existing System:

Cybercrime has undergone a revolutionary change, going from being productoriented to service- oriented because the fact it operates in the virtual world, with different spatial and temporal constraints, differentiates it from other crime taking place in the physical world. As part of this change, the cybercrime underground has emerged as a secret cybercrime marketplace because emerging technological changes have provided organized cybercriminal groups with unprecedented opportunities for exploitation.

The cybercrime underground has a highly professional business model that supports its own underground economy. This business model, known as CaaS, is "a business model used in the underground market where illegal services are provided to help underground buyers conduct cybercrimes, such as attacks, infections, and money laundering in an automated manner,".

Thus, CaaS is referred to as a do-it-for-me service, unlike crimeware which is a do-it-yourself product. Because CaaS is designed for novices, its customers do not need to run a hacking server or have high-level hacking skills.

Consequently, the CaaS business model can involve the following roles: writing a hacking program, performing an attack, commissioning an attack, providing an attack server (infrastructure), and laundering the proceeds.

Sood and Enbody have suggested that crimeware marketplaces have three key elements, namely actors (e.g., coders, operators, or buyers), value chains, and modes of operation (e.g., CaaS, pay-per-install, crimeware toolkits, brokerage, or supplying data). Periodic monitoring and analysis of the content of cybercrime marketplaces could help predict future cyber threats.

3.1 PROPOSED SYSTEM:

The goal of our data analysis framework is to conduct a big-picture investigation of the cybercrime underground by covering all phases of data analysis from the beginning to the end. This framework comprises four steps: (1) defining goals; (2) identifying sources; (3) selecting analytical methods; and (4) implementing an application. Because this study emphasizes the importance of RAT for analyzing the cybercrime underground, the proposed RAT-based definitions are critical to this framework: **Steps 1–4** all contain the RAT elements

A.**Step 1:** Defining Goals The first step is to identify the conceptual scope of the analysis. Specifically, this step identifies the analysis context, namely the objectives and goals. To gain an in-depth understanding of the current CaaS research, we investigated the cybercrime underground, which operates as a closed community. Thus, the goal of the proposed framework is to "investigate the cybercrime underground economy."

B.Step 2: Identifying Sources the second step is to identify the data sources, based on the goals defined by Step 1. This step should consider what data is needed and where it can be obtained. Since the goal of this study is to investigate the cybercrime underground, we consider data on the cybercrime underground community. We therefore collected such data from the community itself and obtained a malware database from a leading global cyber security research firm. Because cybercriminals often change their IP addresses and use anti-crawling scripts to conceal their communications, we used a self-developed crawler that can resolve captcha's and anti-crawling scripts to gather the necessary data. We collected a total of 2,672,091 posts selling CaaS or crimeware, made between August 2008 and October 2017, from a large hacking community site (www.hackforums.net) with over 578,000 members and more than 40 million posts. We also collected 16,172 user profiles of sellers and potential buyers, based on their communication histories, as well as prices and questions and answers about the transactions. The black market uses traditional forum threads (e.g., bulletin boards) instead of typical e-commerce platforms (e.g., eBay, and Amazon). For example, sellers create threads in marketplace forums to sell items, and potential buyers comment on these threads. One of the most significant challenges was therefore converting this unstructured data into structured data.

SYSTEM REQUIREMENTS SPECIFICATION

A **Software Requirements Specification** (**SRS**) is a document that describes the nature of a project, software or application. In simple words, SRS document is a manual of a project provided it is prepared before you kick-start a project/application. This document is also known by the names SRS report, software document. A software document is primarily prepared for a project, software or any kind of application.

There are a set of guidelines to be followed while preparing the software requirement specification document. This includes the purpose, scope, functional and nonfunctional requirements, software and hardware requirements of the project. In addition to this, it also contains the information about environmental conditions required, safety and security requirements, software quality attributes of the project etc.

The system requirement and specification of our project is as follows:

4.1 Function Requirements:

4.1.1 Purpose of the requirements document

A requirements document is a document containing all the requirements to a certain product. It is written to allow people to understand *what* a product should do. Purpose and ((scope, from both a technical and business perspective.Product overview and use cases Requirements, including functional requirements (e.g. what a product should do) usability requirements technical requirements (e.g. security, network, platform, integration, client) environmental requirements support requirements interaction requirements (e.g. how the product should work with other systems) Assumptions Constraints Dependencies

4.2 Non-functional requirements:

4.2.1 Practicality

Project practicality is where you use only what is viewed as least needed to meet your goals. A communication system like tis needs hundreds of thousands of users to survive and thrive. Therefore, it should be designed to support large numbers of users, e.g., a substantial percent of a town or a campus.

4.2.2 Efficiency

Efficiency measures how well and productively a manger uses his resources to achieve

goals. Project management places heavy focus on how to acquire the right project team to perform project tasks and to close project successfully within the agreed constraints.

4.2.3 Cost

Cost management is concerned with the process of planning and controlling the budget of a project or business. It includes activities such as planning, estimating, budgeting, financing, funding, managing, and controlling costs so that the project can be completed within the approved budget. Cost management covers the full life cycle of a project from the initial planning phase towards measuring the actual cost performance and project completion.

4.2.4 Flexibility

The need for flexibility is to deal with changed circumstances. Flexibility is used to scale back activities needing less effort while diverting resources to areas with unexpected problems.

4.2.5 Modularity

Modularity refers to the concept of making multiple <u>modules</u> first and then linking and combining them to form a complete system. Modularity enables re-usability and minimizes duplication. In addition to re-usability, modularity also makes it easierto fix problems as bugs can be traced to specific system modules, thus limiting the scope of detailed error searching

4.2.6 Extensibility

Extensibility is a <u>software engineering</u> and <u>systems design</u> principle where the implementation takes future growth into consideration. The term extensibility can also be seen as a systemic measure of the ability to extend a <u>system</u> and the level of effort required to implement the extension. Extensions can be through the addition of new functionality or through modification of existing functionality.

4.2.7 Reliability

Reliability refers to the probability and or the likelihood that a given product will perform in the way and or manner it was intended to perform in the efforts that have been deemed required of that given product within or under a specific period of time required.

4.2.8 Maintainability

It is the probability that a system or system element can be repaired in a defined environment within a specified period of time. Increased maintainability implies shorter repair times.

4.2.9 Portability

It is a measure of how easily an application can be transferred from one computer environment to another. A computer software application is considered portable to a new environment if the effort required to adapt it to the new environment is within reasonable limits. The phrase "to port" means to modify software and make it adaptable to work on a different computer system.

4.3 SOFTWARE REQUIREMENTS:

Operating system : Windows 10 ,Linux

Coding Language : Python.

Front-End Designing: Python, Html, css, javascript.

Data Base : MySQL.

Domain : Machine Learning and Data Analytics.

4.2 HARDWARE REQUIREMENTS:

System: Intel i3 2.4 GHz.

Hard Disk: Min 100GB.

Monitor: 14' Colour.

Mouse: Optical Mouse.

Ram: Min 2 GB.

4.3 Software Requirements Specification (SRS) Table

In this Chapter we specifically give a brief about Requirement Traceability Matrix which is mainly to check whether the requirements are satisfied or not.

Serial No.	Requirements	Requirement	Requirement Description
	ID	In Brief	
1	RID1	Administrator	admin is an business entiry, person to maintain the entire system.
2	RID2	Users	Users is the one who visits the application and can access the basic information.
3	RID3	System Maintainers (Registered User)	System Maintainer is the service receiver
4	RID4	Home Page	When we run the application we get the application home page.
5	RID5	Login Module	User of the system gets login using the credentials (id an password).
6	RID6	Dataset	In this module we collect opinion dataset from the reputed websites
7	RID7	Data Mining	Process of analysing the data from different perspective and summarize it into useful information.
8	RID8	Machine Learning	Used to develop softwares, microsoft technology more developer friendly.
9	RID9	Django,Pycharm	Used to develop browser based application.
10	RID10	Python	More Suitable programming langauage for real time projects.
11	RID11	MYSQL Server	Microsoft technology used to store the data (opinions data)
12	RID12	Browsers	IE, Google Chrome, Firefox

 Table 4.3: Requirement Traceability Matrix (RTM)

ARCHITECTURE

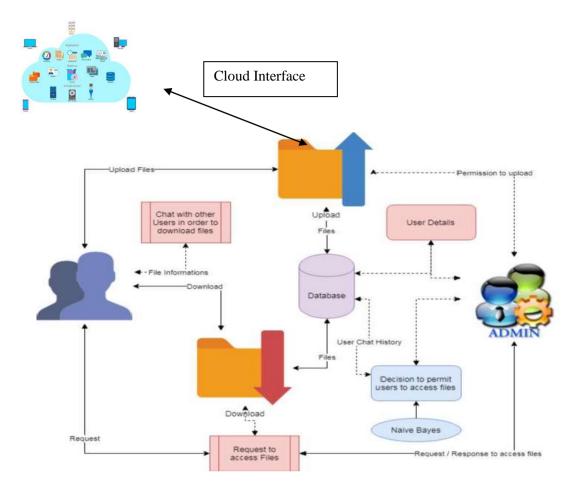


Figure 6.0: Architecture of proposed model.

Step 1:User upload

Users upload their data through any media to the database or cloud server.

- > Step 2:Admin process.:Admin response for the request and process data.
- Step 3:Decisions to provide access
 Using algorithm for decision making and provide access to requested customers.
- > Step 4:Download access files.

Tracking of files and download the requested files.

5.1 MODULES:

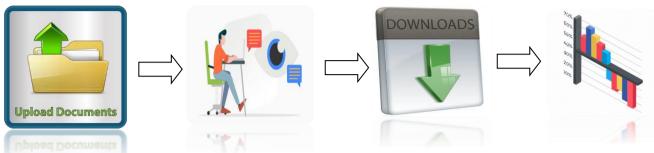


Figure 6.0: Modules for Implementation

Modules that are Implementing this system include the following,

1. Upload Files

Users are allowed to upload the files with the tags given. Once the file is uploaded, then it is sent to approval from admin to publish or make view to other users. These uploaded files can be in any form document, audio or video but not allowed to upload the executable (.exe) files.

2. Conversation Monitoring

Users are allowed to communicate among the other users. This could be monitor by the admin. The malicious conversion likes to threaten the data. In order to protect the cybercrime and prevents from forming cybercrime community. This can be achieved by the help of classification algorithm named naïve Bayes classification.

3.Download Files

The files can be downloading by requesting for the file and once admin approved the files then can be downloadable. The decision to approve files can be taken from the conversation between users. Admin takes the action on download files and approvable status of users. The users are allowed further actions based on the users.

4.Graphical Representations

The analyses of proposed systems are calculated based on the approvals and disapprovals. This can be measured with the help of graphical notations such as pie chart, bar chart and line chart. The data can be given in a dynamical data.

ALGORITHM

6.1Naive Bayes Classifier:

Naive Bayes is a classification algorithm for binary (two-class) and multi-class classification problems. The technique is easiest to understand when described using binary or categorical input values. It is called naive Bayes or idiot Bayes because the calculation of the probabilities for each hypothesis is simplified to make their calculation tractable. Rather than attempting to calculate the values of each attribute value P(d1, d2, d3|h), they are assumed to be conditionally independent given the target value and calculated as P(d1|h) * P(d2|H) and so on. This is a very strong assumption that is most unlikely in real data, i.e. that the attributes do not interact. Nevertheless, the approach performs surprisingly well on data where this assumption does not hold. Make Predictions with a Naive Bayes Model: Given a naive Bayes model, you can make predictions for new data using Bayes theorem.

```
MAP (h) = max(P(d \mid h) * P(h))
```

Using our example above, if we had a new instance with the weather of sunny, we can calculate:

```
go-out = P(weather=sunny | class=go-out) * P(class=go-out) stay-home = P(weather=sunny| class=stay- home) * P(class=stay-home)
```

We can choose the class that has the largest calculated value. We can turn these values into probabilities by normalizing them as follows:

```
P(go-out | weather=sunny) = go-out / (go-out + stay-home) P(stay-home | weather=sunny) = stay-home / (go-out + stay-home)
```

If we had more input variables we could extend the above example. For example, pretend we have a "car" attribute with the values "working" and "broken". We can multiply this probability into the equation. For example below is the calculation for the "go-out" class label with the addition of the car input variable set to "working":

```
go-out = P(weather=sunny | class=go-out) * P(car=working | class=go-out) * P(class=go-out)
```

SYSTEM DESIGN AND IMPLEMENTATION

7.0 DESIGN:

- ➤ In contrast, the goal of our data analysis framework is to conduct a bigpicture investigation of the cybercrime underground by covering all phases of data analysis from the beginning to the end (see Fig. 1). This framework comprises four steps: (1) defining goals; (2) identifying sources; (3) selecting analytical methods; and (4) implementing an application.
- ➤ The data analysis step of the proposed framework involves four steps. Here, we report the data analysis results: CaaS and crimeware classification and market trends, cybercrime market dynamics, and potential hacking targets.
- A. CaaS and Crimeware Classification and Market Trends .
- B. Cybercrime Market Dynamics

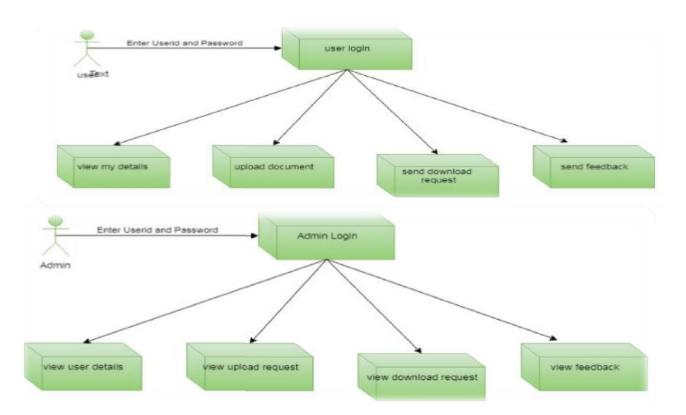
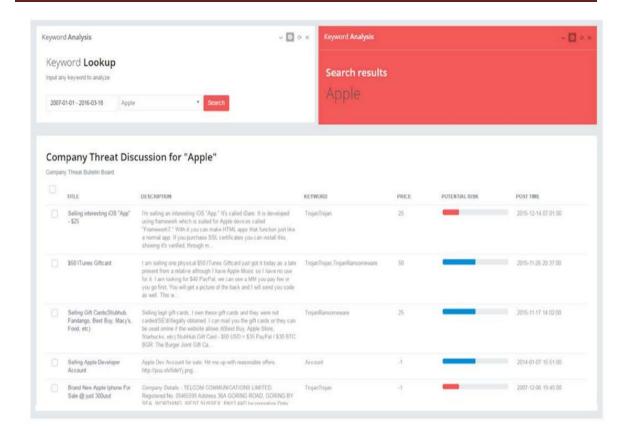


Figure 7.0: System Design Modules.



(a)

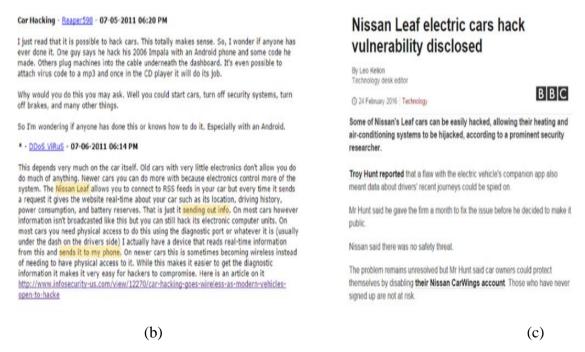


Figure 7.1 Hacking vulnerability disclosed and the earlier signal from the underground: (a) monitoring system. (b)relevant result (July 6, 2011). (c)BBC news (Feb. 24, 2016).