International Journal of Engineering Trends and Technology
ISSN: 2231 – 5381 /doi:10.14445/22315381/IJETT-V68I11P214
Volume 68 Issue 11, 105-112, November 2020

NoSQL Database Design for SNS Profiling in Criminal Investigations

Jiyeon Kim[1,2], Minsun Shim[3], Seyoung Jin[3], Seung Hoon Lee[4], In Soo Lee[4], Myuhng-Joo Kim[2,3,*]

[1]*Center for Software Educational Innovation, Seoul Women's University, Seoul, South Korea*
[2]*Right AI with Security & Ethics(RAISE) Research Center, Seoul Women's University, Seoul, South Korea*
[3]*Dept. of Information Security, Seoul Women's University, Seoul, South Korea*
[4]*Digital Investigations Division, Supreme Prosecutor's Office, Seoul, South Korea*

[1,2]jykim07@swu.ac.kr
[3] {mshim1007, 18sysy}@swu.ac.kr
[4]*{pluviose, insoo}@spo.go.kr*
[2,3,*]mjkim@swu.ac.kr (*Corresponding author)

Abstract — *Social Network Service (SNS) is an online platform for users to communicate and share their interests. As the number of SNS users increases, profiling techniques that collect and analyze user information from SNS have emerged. SNS profiling enables creating personal information of the users and analyzing their interests through their posts, comments, and page likes. SNS profiling has been mainly used as a marketing tool to recommend products through analysis of customer interests. However, SNS profiling has recently been used for various other purposes. In this work, we focus on the use of SNS profiling for criminal investigations. We design a database to store information collected from SNS and propose a model for event profiling. In the database, we create tables of people, events, SNS users, and posts based on NoSQL, a non-relational database, and build a database using DynamoDB. In addition, we conducted a case study for profiling events based on the proposed model.*

Keywords — *SNS profiling, user profiling, event profiling, social network service, open-source intelligence, criminal investigations*

I. INTRODUCTION

The number of Social Network Service (SNS) users has rapidly increased with the spread of smartphones. As of July 2020, the number of global smartphone users is 5.15 billion, with 3.96 billion active social media users [1]. According to Smart Insights, 350 million tweets, 0.83 million Facebook posts, 0.62 million Instagram posts, and 500 h of YouTube videos are being generated in 1 min around the world [2]. SNS profiling has been mainly used as a marketing tool because it allows identification of the user's personal information and their interests by analyzing their posts uploaded in SNS. Numerous studies on social media marketing analyze information obtained from SNS (e.g., user gender, age, location, posts, comments, and page likes) and develop models that recommend customized products after identifying the user preferences [3,4,5]. In the present study, we focus on criminal investigation rather than marketing in connection with the use of SNS data.

By analyzing the data collected from SNS, obtaining the user's personal information and the details of Spatio-temporal activities and user relationships is possible. Such information can also be used as evidence for the investigation. The concept of criminal investigation using SNS was first used in Open Source Intelligence (OSINT). OSINT is intelligence derived from publicly available information collected, utilized, and distributed promptly to appropriate individuals to address a specific intelligence requirement [6]. As the data collected from SNS becomes large, the importance of SNS in OSINT also increases. To use the information obtained from SNS for criminal investigation, to store the collected data in a standardized form, and to develop a theoretical model for a variety of investigational perspectives are necessary. In this study, we design and build a database to store the data collected from SNS and propose a model for event profiling. We focus on the representative SNS such as Facebook and Instagram. We design databases for people, events, SNS users, and posts based on the NoSQL database, a non-relational database. Furthermore, we build these databases using DynamoDB, which is a NoSQL database developed by Amazon. The remainder of this paper is arranged as follows. We review the studies on SNS-based user profiling in Section II. We propose a database design for event profiling using SNS in Section III. We present our implemented database using DynamoDB and our conducted case study for event profiling in Section IV. Finally, the conclusions are presented in Section V.

II. RELATED WORKS

A. *Examples of OSN-based User Profiling*

Shu et al. [7] reviewed user identity linkage's main performance based on Online Social Network (OSN) and proposed a general framework of user identity linkage. The framework consists of two stages: feature extension and model construction. Three features exist profile, content, and network. Besides, the model is implemented in three methods: supervised, semi-supervised, and unsupervised.

The profile feature is a public field of a typical user profile such as username, screen name, location, biography, education, gender, age, occupation, e-mail address, and personal homepage link. The content feature represents the user activities in OSN, such as posting, commenting, and replying. It consists of temporal, spatial, and post (text or image) information. The network feature represents the interactions among the users on a certain OSN and consists of local and global networks based on the network-topology structure's completeness and connectivity.

Malhotra et al. [8] proposed a method on how to uniformly collect, combine, and match the user online digital footprints and suggested a user-profile disambiguation system that could classify the same user's different social networks. The result showed 98% accuracy, 99% precision, and 96% recall using a promising set of features and similar metrics. To measure the similarity between two different social networks, information such as username, display name, description, location, profile image, number of user connections, and context-specific techniques (e.g., Jaro Winkler similarity and Wordnet-based ontologies) were applied. In particular, their studies focused on Twitter and LinkedIn.

Farnadi et al. [9] proposed a hybrid user-profiling architecture in deep neural networks for social-media user profiling. They used three data sources: textual, visual, and relational data sources, from the status updates, profile pictures, and page likes of Facebook users. The framework efficiently integrates sharing and non-sharing representations among different data sources. They extracted the characteristics of social-relational contents and inferred the age, gender, and personality characteristics of social media users for the first time based on Node2Vec embedding. In addition, they experimented with inferring the age, gender, and personality characteristics of Facebook users. In the experiment, they predicted over 50,000 user ages, genders, and personalities. Furthermore, the result demonstrated high accuracy in terms of age and gender predictions.

Vasanthakumar et al. [10] demonstrated the limitations of an extensive data-collection process for OSN user profiling. They proposed a profiling system that applied various approaches based on user activity at a specific site for a certain period. The system consisted of various profiling approaches for each application, such as clustering, face detection, user activities, content analysis, and behavioral analysis. The types of raw data collected by the OSN consisted of activities, images, text data, and geographic location.

We categorized the features used in the references based on where the information is obtained. Table 1 lists these categories, references, and features.

TABLE I
FEATURES USED BY REFERENCE PAPERS IN USER PROFILING

Category	Ref.	Features
Profile	[7]	Username, screen name, location, biography, education, avatar, gender, age, occupation, e-mail, URL
	[8]	Username, display name, description, location, profile
	[9]	Profile picture
Activities and Content	[7]	Posting, commenting, replying: temporal, spatial, post(text or image) information
	[8]	Image
	[9]	Status updates, user page likes
	[10]	Activities, image, text data, geographic location
Network	[7]	Local network, a global network
	[8]	Number of connection

B. Database Design for User Profiling

Sim et al. [11] analyzed mobile data such as the SNS and their application in mobile phones to check whether words related to children, cars, and companion animals were present. They generated a profile to find out the probability of their existence. In the children-related information, they divided it into two relationships: frequently used words and related brands. Each relationship consisted of features created through the selection of words with a frequency of 100 or more.

Tang et al. [12] demonstrated the limitations of existing SNS or Web-based profiling such as profile extraction, name disambiguation, and user interests. By solving the problems, they proposed a model that provided better Web services. They also designed a database schema by extending the friend-of-a-friend ontology. Four relationships exist, researcher, publication, interest, and social network. The social network contains two features: researchers and relationship. They are connected to the personalized_social relationship. The researcher relationship contains features such as name, homepage, personal photo, and position. The publication relationship contains features such as title, date, and download_url. These two relationships are connected by authored_by and author relationships. The interesting relationship contains features such as the topical aspect and time, and these two features are connected by has_interest. The social-network relationship has two features: researchers and relationship, and these two are connected by personalized_social.

Pellet et al. [13] proposed a module and a process for profiling or guessing the location based on the activities of SNS users. As SNSs, Facebook, Instagram, and Twitter were used to gather the information; the information is stored in a database consisting of five relationships: posts, users, relations, locations, and location history. Posts (which indicates what post was posted), relations (which indicates what type of relationship exists among the users), and LocationHistory (which indicates what SNS the user has visited) are each connected to the user relationship. Locations (which indicates information about a certain place) and LocationHistory are connected. In each relationship, the related information can be stored in the related features. Table 2 lists the relational databases and schemas of the references mentioned above.

TABLE II
RELATIONSHIPS OF THE SCHEMA AND FEATURES IN THE DATABASE

Ref.	Relationships	Features
[11]	Frequently used words	Children, kids, school, son, baby, mother, father, etc.
	Related brands	Lego, AramBooks, micro kickboard, etc.
[12]	Researcher	Name, homepage, person photo, address, phone, etc.
	Publication	Title, publication_venue, start_page, end_page, etc.
	Interest	Topical aspect, time
	Social network	Researchers, relationship
[13]	Users	idUser, socialPlatform, targetedFirst, nameUser
	Posts	idPost, txt, time
	Relations	idRelation, wasWithCount, shareCount, metionsCount, etc.
	Locations	Id Location, lat, lng, placename, src
	LocationHistory	idLHistory, time

III. DESIGN OF DATABASE FOR EVENT PROFILING BASED ON SNS

To investigate a criminal case, a "case" database and a "person" database containing information about the main people and their connections are needed. In this work, we design these two databases plus two more new databases, namely, "SNS Users" and "SNS Posts," because the events are analyzed through SNS profiling.

A. *Event and User Databases*

The event database contains StartTime, EndTime, StartDate, EndDate, location fields, and people field, indicating the people related to the event. Furthermore, it contains the description and reference fields. Table 3 lists the field names, their types, and their descriptions in the event database.

TABLE III.
FIELDS IN THE EVENT DATABASE

Field	Type	Description
Id	String	Case number
StartTime	String	The start time of the case
StartDate	String	Start date of the case
EndTime	String	The end time of the case
EndDate	String	The end date of the case
Location	String	The location where the case happened
People	String Set	People involved in the case
Description	String	Description of the case
Reference	String Set	Source of the case

The user database consists of name, age, birth, and education, which are verified through profiling. The person database is separately made from the SNS user database because people often create social media accounts with inaccurate or distorted information. Table 4 lists the field names, their types, and their descriptions in the User database.

TABLE IV.
FIELDS IN THE USER DATABASE

Field	Type	Description
Id	String	User number
Name	String	Name of the user
Birth	String	The date when the user was born
Age	Number	The age of the user
E-mail	String	E-mail address of the user
Phone	String	Phone number of the user
Job	String	Where the user works
Hobby	String Set	User's hobby
Family	String Set	Id numbers of the family of the user in the User database
Education	String Set	User's educational background (e.g., university, high school)
Facebook	String	User's Facebook account id number in the Facebook database
Instagram	String	User's Instagram account id number in the Instagram database

B. *SNS Users and Post Databases*

We designed databases to store user-profiles and posts as follows.

a) Facebook User Profile: Through the Facebook database, we can collect user information and their human network data, such as their friends and family. Some features have overlapped information with Table 3. However, because the SNS user profile is based on the user's input data, the real data in the feature can be different. The user profile collected from Facebook can be stored in Table 5.

TABLE V.
FIELDS IN THE FACEBOOK DATABASE

Field	Type	Description
Id	String	User number
Username	String	Name of the user
Profile_Pic	String Set	Profile photo
Back_Pic	String Set	Background photo
Workplace	String Set	Where the user works
University	String	The university the user attends or graduated
Major	String	The major of the user
Schools	String Set	Graduated schools of the user
Residence	String	Where the user lives
native place	String	Where the user lived
Address	String	The address of the user
Phone	String	The phone number of the user

Extended_ URLs	String Set	The social, website link, or personal blog of the user
E-mail	String	E-mail address of the user
Birth	String	The date when the user was born
Age	Number	The age of the user
Family	String Set	Id numbers of the family of the user in the Facebook database
Lover	String	The lover of the user
Gender	String	The gender of the user
Religion	String	The religion of the user
Politics	String	The political view of the user
Description	String	The personal description
NickName	String	The nickname of the user
Favorite Phrase	String	The favorite phrase of the user
main events	String Set	The main event selected by the user
Friends	String Set	Id numbers of the friends of the user in the Facebook database
PageLikes	String Set	The liked pages of the user
CheckIN_ Place	String Set	The check-in place of the user
CheckIN_ Date	String Set	The check-in date of the user
Post	String Set	The uploaded posts of the user
Tagged_ Post	String Set	The posts that the user gets tagged in
Instagram _story	String Set	The story posted on Instagram
Profile_URL	String	The URL of the profile

b) Facebook Post: Through content, such as photos and videos, location, hashtags, and comments in the post, one can obtain temporal and spatial information of the user. Additionally, through comments from the posts, shared posts, page likes, main events, and check-in information, one can identify the user's interests. The data related to posts collected from Facebook can be stored in Table 6.

TABLE VI.
FIELDS IN THE FACEBOOK POST DATABASE

Field	Type	Description
Post_Id	String	Post number
Post_URL	String	URL of the post
Original_Post	String	URL of the original post
Uploader	String	The user who wrote the post
Upload_Date	String	The date when the post uploaded
Upload_Time	String	The time when the post uploaded
Participant	String Set	The participants of the post
Place	String	The place of the post
Text	String	The content of the post
Photo	String Set	The photos in the post
Video	String Set	The videos in the post
HashTag	String Set	The hashtag of the post
Comment_ Group	String	The group number of the comment
Comment_Time	String	The time of the comment
Comment_User	String	The user who wrote a comment
Comment_Text	String	The content of the comment
Comment_Like	String Set	The user who 'liked' the comment
Post_Like	String Set	The user who 'liked' the post
Post_Best	String Set	The user who 'loved' the post
Post_Cheerup	String Set	The user who 'cared' the post
Post_Funny	String Set	The user who 'haha' the post
Post_Cool	String Set	The user who 'wowed' the post
Post_Sad	String Set	The user who 'sad' the post
Post_Angry	String Set	The user who 'angry' the post

c) Instagram User Profile: The information collected from Instagram is as in Table 7.

TABLE VII.
FIELDS IN THE INSTAGRAM DATABASE

Field	Type	Description
Id	String	User number
UserID	String	The id of the user
Username	String	The name of the user
Profile_Pic	String	The profile photo of the user
Description	String	The description of the user
University	String	The university the user graduated
Major	String	The major of the user
Extended_ URLs	String Set	The social, website link, or personal blog of the user
Birth	String	The date when the user was born
Age	Number	The age of the user
Family	String Set	Id numbers of the family of the user in the Instagram database
Lover	String	The user's partner
Gender	String	The gender of the user
Follower	String Set	The list of followers
Following	String Set	The list of followings
HashTag_ Following	String Set	The list of the following hashtags
Post	String Set	The list of the post
Tagged_ Post	String Set	The posts that the user gets tagged in
Highlight_ Story	String Set	The highlight story posts
Story	String Set	The story posts
Profile_URL	String	The URL of the profile

Compared to the Facebook database, there are no Back_Pic, Workplace, Schools, Residence, NativePlace, Address, Phone, E-mail, Religion, Politics, Nickname, FavoritePhrase, MainEvents, Friends, PageLikes, CheckIN_Place, and CheckIN_Date fields in the Instagram database. Additionally, UserId, Follower, Following, HashTag_Following, Highlight_Story, and Story fields were added. In particular, the Story field is deleted automatically within 24 hours.

d) Instagram Post: Through uploaded date, location, content, photos, videos, hashtag, comments, and story information, one can obtain the temporal and spatial information of the user. Additionally, through the story, highlight story, following lists, hashtag following lists, and hashtag information, one can identify the user's interests. The information collected from Instagram posts is as in Table 8.

TABLE VIII.
FIELDS IN THE INSTAGRAM POST DATABASE

Field	Type	Description
Post_Id	String	Post number
Post_URL	String	URL of the post
Uploader	String	The user who wrote the post
Upload_Date	String	The date when the post uploaded
Participant	String Set	The participants of the post
Place	String	The place of the post
Place_NewPost	String Set	The latest post based on the location
Place_HotPost	String Set	The most popular post based on the location
Text	String	The content of the post
Photo	String Set	The photos in the post
Video	String Set	The videos in the post
HashTag	String Set	The hashtag of the post
HashTaag_ NewPost	String Set	The latest post based on the hashtag
HashTag_ HotPost	String Set	The most popular post based on the hashtag
Comment_ Group	String	The group number of the comment
Comment_Time	String	The time of the comment
Comment_User	String	The user who wrote a comment
Comment_Text	String	The content of the comment
Comment_Like	String Set	The user who 'liked' the comment
Post_Like	String Set	The user who 'liked' the post

Compared to the Facebook Post database, there are no Original_Post, Upload_Time, Post_Best, Post_Cheerup, Post_Funny, Post_Cool, Post_Sad, and Post_Angry fields in the Instagram database. In addition, UserID, Place_NewPost, Place_HotPost, HashTag_NewPost, and HashTag_HotPost fields were added in the Instagram database. The location, most popular posts based on hashtag, and latest posts based on a hashtag can be additionally collected. However, the uploaded time data of the post could not be identified.

IV. SNS PROFILING USING NOSQL DATABASE

Based on the structure we designed in Section 3, we developed a database using DynamoDB, which is the NoSQL database. Events, Users, SNS Users, and posts are each used to form a table. For the case study, we made tuples of data, as shown in Figures 1–4. Figure 1 shows a tuple in the Event table.


```
"TableData": [
  {
    "ID": "E1"
     "StartDate": "11/12/2020"
     "StartTime": "15:00"
     "EndDate": "11/12/2020"
     "EndTime": "18:00"
     "Location": "U1's home"
     "People": ["U1","U2"]
     "Description": "Assault and theft."
     "Reference": ["FP1"]
  }
]
```


Fig. 1 Tuple example in the Event table (Id: E1)

Figure 1 contains data related to event E1. Through the StartDate and StartTime fields' data, the start time of event E1 can be determined. Similarly, the end time of the event was identified by using the EndDate and EndTime fields. Through the Location and People fields, it is found that the event occurred in the home of U1 with U1 and U2 as the accomplices. Through the Description field, it is observed that the event was assault and theft. Additionally, the source of the event was FP1 (Facebook Post 1) through the Reference field.


```
{
  "ID": "U1"
  "Name": "Bella"
  "Birth": "05/12/1995"
  "Age": "26"
  "Email": bella@gmail.com
  "Phone":"+82-10-1234-5678"
  "Job": "Student"
  "Hobby":["Music","Painting"]
  "Family": ["Mother","Father"
            ,"Younger Sister"]
  "Education":["Seoul Women's
              University"]
  "Facebook": "F1"
  "Instagram": "I1"
}
```

```
{
  "ID":"U2"
  "Name":"Amy"
  "Birth":"10/30/1995"
  "Age": "26"
  "Email": "amy@gmail.com"
  "Phone":"+82-10-8765-4321"
  "Job": "Student"
  "Hobby": ["Sing","Dancing"]
  "Family": ["Mother","Father"
            ,"Sister"]
  "Education":["XXhighschool"
              ,"Seoul Women's
               University"]
  "Facebook": "F2"
  "Instagram": "I2"
}
```


Fig. 2 User tuple related to E1 (Id: U1, U2)

The user table shown in Figure 2 contains the general attributes gained from profiling the user. U1, whose name is Bella and was born on October 30th, 1995, is one of the accomplices. Moreover, information, such as e-mail, phone number, job, family, and education about U1 and U2 can be collected. Because the Age and Education data of U1 and U2 are similar, it can be said that they are university friends, and their Facebook and Instagram accounts can be identified. Figure 3 shows a tuple in the Facebook Post table.

```
{
  "Post_ID":"FP1"
   "Post_URL": "https://exampleFP1.com"
   "Uploader": "F1"
   "Upload_Date":"11/14/2020"
   "Upload_Time":"13:00"
   "Participant":"F2"
   "Place": "F1's home"
   "Text": "Amy came to my house on November 12th.
   Amy assaulted me and stole my watch."
   "Photo": "Photo3"
   "Video": "Video1"
   "Comment_Group": "1"
   "Comment_Time": "11/14/2020"
   "Comment_User": "F3"
   "Comment_Text": "Are you sure? Are you okay?"
   "Comment_Like": ["F3"]
   "Post_Like": ["F3"]
   "Post_Cheerup": ["F3"]
   "Post_Sad": ["F3"]
   "Post_Angry": ["F3","F4"]
}
```


Fig. 3 Facebook Post tuple related to E1 (Id: FP1)

According to the source data extracted from the Reference field in E1, we can find the post on Facebook. By obtaining the data from the Facebook field of U1, we can observe that the Uploader field data in FP1 is U1. According to the information collected from the Facebook field of U2, the accomplice of the post FP1 is F2. Comparing the event's occurred time from the E1 tuple and the Upload_Data and Upload_Time of the FP1, it is evident that the post was uploaded after the event. Furthermore, by the context of the Text field, we can obtain more descriptions about the event. From this information, U2 (similar to F2) assaulted U1 (similar to F1) in the house of U1 and stole the watch. Additionally, related information, such as Photo3 and Video 1, may be obtained.

Figure 4 shows a tuple in the Facebook user table, and Figure 5 shows a tuple in the Instagram user table. Using the data collected from the Facebook and Instagram profile of U2, we can verify whether the information is correct or incorrect. Furthermore, from Figure 5, we can gather more information by collecting more items that were not in the user table of the U2 tuple.


```
{
  "ID": "F2"
   "Username": "Amy"
   "Profile_Pic": "Profile_Pic2"
   "Back_Pic":"Back_Pic2"
   "Workplace":"Workplace2"
   "University":"Seoul Women's University"
   "Major": "Dept. Information Security"
   "Schools":["Example HighSchool","Example Middle
                School"]
   "Residence": "Seoul"
   "NativePlace": "Korea,Seoul"
   "Address": "Seoul-Itewon"
   "Phone": "+82-10-8765-4321"
   "Extended_URLs":"www.ex-twitter1.com"
   "Email": "Amy@gmail.com"
   "Birth": "10/30/1995"
   "Age": "26"
   "Family": ["F10","F11","F12"]
   "Lover": "None"
   "Gender": "F"
   "Religion": "None"
   "Politics": "None
   "Description": "Hi, I'm Amy"
   "MainEvents": "The student body president"
   "Friends": ["F1","F3","F4","F5"]
   "PageLikes": "Seoul Women's University"
   "CheckIN_Place": "N Seoul Tower"
   "CheckIN_Date": "12/24/2019"
   "Post": ["FP2","FP3","FP4"]
   "Instagram_stroy": ["IP5"]
   "Profile_URL": "https://www.amy_fb.com"
}
```


Fig. 4 Facebook user tuple related to E1 (Id: F2)


```
{
   "ID": "I2"
   "UserID": "@Insta_Amy"
   "Username": "Amy"
   "Profile_Pic": "Profile_Pic4"
   "Description":  "hello, i'm amy"
   "University": "Seoul Women's University"
   "Major": "Dept. Information Security"
   "Extended_URLs": "www.ex-twitter1.com"
   "Birth": "10/30/1995"
   "Age":  "26"
   "Family": ["I10","I11","I12"]
   "Lover": "None"
   "Gender": "F"
   "Follower": ["I1","I3"]
   "Following": ["I1","I3"]
   "Post": ["IP2","IP3","IP4"]
   "Story":  "IP5"
   "Profile_URL": "https://www.amy_insta.com"
}
```


Fig. 5 Instagram user tuple related to E1 (Id: I2)

Figure 6 shows the overall table structure, which saves the Facebook posts of the user. By setting PartitionKey to the Uploader and sortkey to the Upload_date of Global Secondary Indexes (GSI), we can re-sort based on the uploaded data.


```
"GlobalSecondaryIndexes": [ {
   "IndexName": "Uploader",
   "KeyAttributes": {
      "PartitionKey": {"AttributeName": "Uploader"},
      "SortKey":{"AttributeName": "Upload_Date" },
    },
    "Projection": { "ProjectionType": "ALL" }
}]
```


Fig. 6 Settings and structure about the GSI of the Facebook post table

The process of profiling case E1 is as shown in Figures 7–10. Profiling obtains the tracks of the accomplice using SNS. Therefore, from the Facebook post of U2, we should gain the alibi when E1 occurred.

```
Data Plane operations: Query
Table: User
Partition Key: U2
Projection expression: Instagram, Facebook
```

```
aws dynamodb query
   --table-name User
    --key '{"Id": {"U2"}}'
   --condition-expression
   "attribute_exists(Instagram,facebook)"
```

Fig. 7 Query and pseudocode to obtain the Facebook and Instagram account of U2 from the user table

#	Instagram	Facebook
1	I2	F2

Fig. 8 Result tuple of the query in Figure 7

Through the query from Figure 7, the SNS account data of U2, who is the accomplice of case E1, can be gained. The Instagram id number of the U2 was I2, and the Facebook id number was F2.

```
Data Plane operations: Query
Table: Facebook_Profile
Partition Key: F2
```

```
aws dynamodb get-item --consistent-read
   --table-name Facebook_Profile
   --key '{ "ID": {"F2"}}'
```

Fig. 9 Query and pseudocode to obtain the Facebook account of U2 from the Facebook table

#	Profile_Pic	Post	Profile_URL	Username
1	["Profile_Pic2"]	["FP2","FP3","FP4"]	https://www.amy_fb.com	Amy

Fig. 10 Result tuple of the query in Figure 9

Through the query from Figure 9, from the Facebook_Profile table, the profile of F2 can be gained. Figure 10 shows Profile_URL, Profile_Pic, Username, and Post fields of 33 fields.

```
Data Plane operations: Scan
Table: Facebook_Post
Secondary index: No index chosen
Filter expression: AND
Condition: “Uploader = F2”
Condition: “Upload_Date = 11/12/2020”
```

```
aws dynamodb scan
   --table-name Facebook_Post
    --scan-filter '{
       "Uploader": {
          ComparisonOperator: "EQ",
          AttributeValueList: [ {"F2"} ]
       },
    "Upload_Date": {
       ComparisonOperator: "EQ",
       AttributeValueList: [{"11/12/2020"}]
    }
}'
```

Fig. 11 Query and pseudocode to obtain the Facebook post of F2 when the case occurred

#	Post_ID	Post_URL	HashTag	Place	Participant
1	FP3	https://exampleFP...	["movie"]	F1's home	["F1"]

Text	Photo	Upload_Time
Watching a movie at Amy's house after a long time!	["Photo5"]	14:28

Fig. 12 Result tuple of the query in Figure 11 (Post_ID: FP3)

```
{
   "Post_ID": { "S": "FP3"},
    "Post_URL": { "S": "https://exampleFP3.com"},
    "Uploader": {"S": "F2"},
    "Upload_Date" : {"S": "11/12/2020"},
    "Upload_Time" : { "S": "14:28"},
    "Participant" : {"SS": ["F1"] },
    "Place": {"S": "F1's home"},
    "Text": {"S": "Watching a movie at Amy's house
                  after a long time!"},
    "Photo": { "SS": ["Photo5"},
    "HashTag": {"SS": ["movie"},
    "Post_Like": { "SS": ["F1"]},
    "Post_Best": {"SS": ["F1"]},
}
```

Fig. 13 Json format of the result in Figure 12

The FP3 post of U2, which was uploaded between when the case occurred, is gained as in Figure 12 through the query in Figure 11. F2 uploaded the photo of watching a movie with F1 at the house of F1. Therefore, we can find out that U2 was at the place where the case occurred during that time.

```
Data Plane operations: Query
Table: Instagram_Profile
Partition Key: I2
```

```
aws dynamodb get-item
   --consistent-read
   --table-name Instagram_Profile
   --key '{ "ID": {"I2"}}'
```

Fig. 14 Query and pseudocode to obtain the Instagram profile of U2 from Instagram_Profile table

#	Profile_Pic	Post	Profile_URL	Username
1	Profile_Pic4	["IP2","IP3","IP4"]	https://www.amy_insta.com	Amy

Fig. 15 Result tuple of the query in Figure 14

Through the query from Figure 14, from the Instagram_Profile table, the profile of I2 can be gained. Figure 15 shows Profile_URL, Profile_Pic, Username, and Post fields of 21 fields.

```
Data Plane operations: Scan
Table : Instagram_Post
Secondary index: No index chosen
Filter expression: AND
Condition: “Uploader = F2”
Condition: “Upload_Date = 11/12/2020”
```

```
aws dynamodb scan
   --table-name Facebook_Post
   --scan-filter '{
     "Uploader": {
       ComparisonOperator:" EQ",
       AttributeValueList: [{”F2”}] },
    "Upload_Date" :{
     ComparisonOperator:" EQ",
```

```
    AttributeValueList: [{"11/12/2020"}]
  }
}'
```

Fig. 16 Query and pseudocode to obtain the Instagram post of I2 from Instagram post table

#	Post_ID	Post_URL	HashTag	Place	Participant
1	IP3	https://exampleIP...	["NEW"]	amy's home	["-"]

Text	Video	Comment_User	Comment_Text
a new wristwatch! :)	["Video4"]	["I4"]	It looks exactly like Bella's watch!

Fig. 17 Result tuple of the query in Figure 16


```
{
  "Post_ID": {  "S": "IP3"},
   "Post_URL": {"S": "https://exampleIP3.com"},
   "Uploader": { "S": "I2"},
   "Upload_Date": {"S": "11/13/2020"},
   "Participant": {"SS": [ "-"]},
   "Place": { "S": "amy's home"},
   "Place_NewPost": {{"SS": [ "-"]},
   "Place_HotPost": {"SS": [ "-"]},
   "Text": {"S": "a new wristwatch! :)"},
   "Photo": {"SS": [ "-"]},
   "Video": {"SS": ["Video4"]},
   "HashTag": { "SS": [ "NEW"]},
   "Comment_Time": {"S": "11/13/2020"},
   "Comment_User": {"S": "[\"I4\"]"},
   "Comment_Text": {"S": "It looks exactly like
                                  Bella's watch!"}
}
```


Fig. 18 Json format of the result in Figure 16

Through the query in Figure 16, the post (Post_ID: IP3, Uploader: I2) of U1 was uploaded the day after the case occurred (Upload_Date: 11/13/2020) is found as in Figure 17. By observing the data in the Text, hashtag, and Comment_text fields, we can find out the new watch, which is similar to the watch of U1. We can conclude that the new watch of U2 is stolen based on the FP1 post uploaded by U1. Thus, we can obtain more information for profiling through friends' posts and comments when the case occurred.

V. CONCLUSIONS

In this study, we present a database model for investigating crime through SNS profiling. By analyzing the profiles and activities of SNS users, one can obtain personal information, such as gender, residence, and occupation, and information, such as the interests and personal contacts of the user. We design a database to analyze various information collected from Facebook and Instagram, which is essential for criminal profiling. We created the incident, the people related to the incident, their SNS profile, and activity information using the NoSQL database, and built it with DynamoDB, the NoSQL database of Amazon. Additionally, by conducting a case study to solve the case through SNS profiling, it was further verified that the proposed database design was effective for criminal profiling. In the future, we plan to improve and expand the database. Data collected from other types of SNS, such as Twitter and Snapchat, can be employed for criminal investigations.

REFERENCES

[1] Smart Insights website. [Online]. Available: https://www.smartinsights.com/social-media-marketing/ social-media-strategy/new-global-social-media-research/ (2020)

[2] Smart Insights website. [Online]. Available: https://www.smartinsights.com/internet-marketing-statistics/happens-online-60-seconds/(2020)

[3] J. Dam and M. Velden, Online profiling and clustering of Facebook users, Decision Support Systems, 70 (2015) 60-72.

[4] A. Antoniou, I. Lykourentzou, J. Rompa, E. Tobias, G. Lepouras, C. Vassilakis, and Y. Naudet. User profiling: Towards a Facebook game that reveals cognitive style. in Proc. International Conference on Games and Learning Alliance : 2013. 349-353.

[5] Shen, B., and Bissell, K., Social media, social me: A content analysis of beauty companies' use of Facebook in marketing and branding, Journal of Promotion Management, 19(5) (2013) 629-651.

[6] Homeland Security Digital Library website. [Online]. Available: https://www.hsdl.org/c/us-national-intelligence-an-overview-2011/ (2012)

[7] Shu, K., Wang, S., Tang, J., Zafarani, R. and Liu, H., User identity linkage across online social networks: A review, ACM Sigkdd Explorations Newsletter, 18(2), (2017) 5-17.

[8] Malhotra, A., Totti, L., Meira Jr, W., Kumaraguru, P. and Almeida, V, Studying user footprints in different online social networks, in Proc. IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining : 2012. 1065-1070.

[9] Farnadi, G., Tang, J., De Cock, M. and Moens, M. F, User profiling through deep multimodal fusion, in Proc. the Eleventh ACM International Conference Web Search and Data Mining, pp. 171-179, 2018.

[10] Vasanthakumar, G. U., Sunithamma, K., Shenoy, P. D. and Venugopal, K. R., An overview on user profiling in online social networks, International Journal of Applied Information System, 11(8) (2017) 25-42.

[11] Myo-Seop Sim and Heui-Seok Lim, Predicting User Profile based on user behaviors, Journal of the Korea Convergence Society, 11(7) (2020) 1-7.

[12] Tang, J., Yao, L., Zhang, D. and Zhang, J, A combination approach to web user profiling, ACM Transactions on Knowledge Discovery from Data, 5(1) , 2010 1-44.

[13] Pellet, H., Shiaeles, S. and Stavrou, S, Localising social network users and profiling their movement, Computers and Security, 81 (2019) 49-57.