Snoogle: A Search Engine for real World

**Parth Patel (ru7853)**

**MS Computer Science**

**California State University, Eastbay**

# **Abstract**

Nowadays, objects like coffee mug toasters creates the wireless network using embedding small device. This embedding small device contain the information of the object. In this paper, I have present Snoogle a search engine for such network. Here user can search particular object that fit the Snoogle description. It uses information retrieval techniques to index data and perform user query. Snoogle also use bloom filter to reduce the communication overhead. Security is built for secure the information about object.

# Introduction:

Embedding small device into everyday object allow us to retrieve and store information about the object. For example, consider there are collection of document binders. Each Binder is connected with small digital device that contain the information of that binders. The information of the binders can be created through input device such as digital pen. The digital pan translate the user word onto text. Whenever user want to find particular document, user query binder’s embedding device to study the contact of the binder and retrieve the binder’s information.

Information of binder should be installed on binder itself not on the other object itself or remote server. This is because of short range improve the robustness and due to short range binder can be also accessed via Bluetooth. Moreover, security and privacy also require for protection data information. Snoogle allows the user to search for information in computing environment. I assume that small devices are already embedded in objects, and each device has limited communication, storage and processing ability. I also assume that effective data input mechanism is used for storing the necessary data into device.

The paper is presented Snoogle system components, system architecture, query algorithms, mobility and security, performance evaluation and system limitation.

# System Design

* 1. System Components:

Snoogle is consist of Index Point (IP), Key Index Point (KeyIP) and object sensors. Object sensor contain the textual information about object. The information of the object is given by the user. IP is associate with object location, for example particular room in school. Gathering information from object and maintaining information of object is possible due to IP. IP have large storage capacity. Collection of IPs create huge network. While KeyIP collect data from particular IP.

* 1. System Architecture:

Snoogle work on two-tier hierarchical architecture. The lower tier consist of index point and object sensor. Index Point manager object sensor with in it range. Here, Object sensor register itself and transfer its information to specific IP. Base on the object information Index Point create index for local search. In upper tier, Index point pass object information to Key Index point and KeyIP maintain that IP’s information. Therefor KeyIP return IP based on user query. Moreover, IP provide message routing for check traffic between IPs, KeyIP and users. KeyIP is collection of IPs and IP is collection of object sensor.

When users want to find any particular object then they can query either local query or distributed query. User query local query on particular IP when they know the specific location of the object. While distributed query is on KeyIP when user does not know the IP, so KeyIP provide some IP which match user query best. After that user perform local query on IP which are provided by KeyIP.

* 1. Data Processing in Object Sensor:

Each object sensor contains of payload data and metadata. Where payload data contain the sort description of connected object. Metadata contain main information about connected object and it is representation of the payload. For example, the payload of an object sensor is connected to folder which contain the sort information about it object. The metadata contain {term1: freq1: id1} ... {termN: freqN: idN}, where term keyword is describe the payload. While freq keyword describe this term that describing the payload. User store the information into object sensor and which is responsible for sending the payload and metadata.

* 1. Data Processing and storage at IPs:

IPs have two data processing roles. User first save the object information on object sensor. Then IPs collect information from object sensor with in their area and store that information into inverted index. For reliability and save space, inverted index is store on board flash memory of sensor. IPs rapidly send aggregated update information to KeyIP so KeyIP can contain accurate information. Index Point perform three operation for maintain object information.

*Insert*: Insert operation perform when any new object sensor send signal to IP. Therefore IP store the new information and object id into table.

*Delete*: Delete operation perform when any object leave particular area of Index Point. For example, a user change the location of coffee mug so object sensor of coffee mug is no longer with in specific IP, at that time IP perform delete operation for remove the information about that object.

*Modify*: Modify operation is combination of the delete and insert operation. Modify operation occur when any object information is change.

# Performing Query:

There is two way that user use to querying Snoogle. The first is query on specific IP. Second is query the KeyIP first, this is distributed query and that return list of IPs. The first query method is used when user want to find an item with in specific location or user know that item is within that specific area. For instance, user want to find pen and he have some idea that pen is within this area. Therefore, he query IP by providing some keyword about pen. Thus, IP gives relevant answers based on user query. Each answer is consist of id of the object. Then user can find object physically or query the sensor.

Another query is used when user want to find object but he does not know the specific location of the object or he has no idea of where is the object. At this time user first query to KeyIP first by providing some keyword that describe the object. Based on keyword KeyIP give ranked list of *m* IPs that contain the objects that best match. Here *m* is system parameter. Then user perform top-k query from the returned *m* IPs to find the object.

# Mobility and Security:

* 1. Security and Privacy:

Snoogle can reveal all personal information to other. For example, user may not want to his private document to be searchable by any unknown person. Therefore Snoogle must have a security mechanism to prevent accessing data from unknown user. In short, user first need to authenticate itself before searching anything. Here Snoogle use public key cryptography to have a clean user interface and simple key management. Snoogle provide security protection by adding a security tag on specific object. Security tag contain OwnerID and GroupMask field. OwnerID contain which user is authorize to access data. GroupMask contain which group of user have privilege to access the object.

If user want to search private object then he first sends query and certificate, where certificate is authorize by Snoogle administration. IP first verify certificate and make sure the corresponding OwnerID and GroupMask match with the object security tag. Then IP uses public key to encrypt a random chosen secret key and send cipher text to user. If user will successfully decrypt cipher text then it allow that user to use that object information.

* 1. Supporting Mobile Objects:

Snoogle uses beacon and timer method to maintain accurate information of IPs. In beacon method IP send rapidly broadcast a beacon message that identified itself. Object sensor within area receive area receive that beacon message and compare it with previous beacon message. If new beacon message match with previous beacon message then it indicate that object is within the same IP. Therefor sensor do nothing. Otherwise sensor will report its metadata and id information to new IP.

In timer method the communication is start by each sensor. Each sensor rapidly send “keep\_alive” message to IP. At that duration, IP maintain timer. If IP does not get “keep\_alive” message from certain object within specific time then IP will consider that that object is not with in their range so IP will delete the information about that object. If IP will receive the “keep\_alive” message within specific time then it will do nothing. Beacon and timer is like “push” and “pull”. In beacon method, IPs pull information from each object sensor. While in timer method, object sensors push their status to IP.

Beacon method consumes less energy than timer method because object sensor only have to wake up in for specific time to listen the beacons. While timer method is more reliable because when object move from one IP to another IP, the old IP get notification of missing using timer method and new IP also get notification that new object arrive in that area using timer message send by the object. Therefore, beacon message is more reliable when object is more stable and timer method is more reliable when object is moving frequently.

# System Limitation:

* 1. Communication Reliability:

Dropping message affect that accuracy of connected object. Due to drop message IP consider that particular object is out of range but actually this is due to message that has been dropped. This is also occur when any object change it location and new IP does not detect moving object due to packet loss during beacon sending and reply. It suggest that more reliable communication might be useful.

* 1. System Scalability:

Consider that KeyIP is consist of to many IP and each IP is consist of to many information of object sensor. Single IP is insufficient when there is too many object. Therefore, Snoogle might be used more power full IP that can store more byte of data.

* 1. Mobility Support:

Consider that an object is keep moving from one location to another location so it will be hard to keep the accurate information about that object. However once object is stop moving for certain time then Snoogle can get the location of that object.

# Conclusion:

In this paper, I have present Snoogle, information retrieval system built on object sensor networks. I introduced query processing to query the data to find location. I also introduce flexible security mechanism and limitation of the system.

# References:

1. J. Li, B. T. Loo, J. M. Hellerstein, M. F. Kaashoek, D. Karger, and R. Morris, “On the feasibility of peer-to-peer web indexing and search” in IPTPS03.
2. H. Wang and Q. Li, ““Efficient Implementation of Public Key Cryptosystems on Mote Sensors (Short Paper)” in International Conference on Information and Communication Security (ICICS), LNCS 4307, Raleigh, NC, Dec. 2006, pp. 519–528.
3. Haodong Wang, Chiu C. Tan, and Qun Li, “Snoogle: A Search engine for pervasive environments” in Parallel and Distributed System, IEEE Transaction, 2010, Volume 21.8, page 1188 – 1202.
4. Haodong Wang, Tan C. C., Qun Li, “Snoogle: A Search Engine for the physical World”, INFOCOM 2008, The 27th Conference on Computer Communication IEEE (2008).
5. B. H. Bloom, “Space/time trade –offs in hash coding with allowable errors”, in Communication of the ACM, 13(7):422-426, 1970
6. S. R. Madden, M. J. Franklin, J. M. Hellerstein, and W. Hong, “TinyDB: an acquisitional query processing system for sensor networks”, ACM Trans, Database System, (2005)