Indian Journal of Emerging Electronics in Computer Communications
Vol.2, Issue 1, 2015, Page.311-317, ISSN: 2393-8366

The Evaluation of Bloom Filter Based Publish/Subscribe System for HUNETs

[1]*K.Sugasini* [2]*V.Prakasham*
[1] *PG Scholar* [2] *Assistant Professor*
[1,2]*Pallavan College of Engineering*

Abstract --To achieve interoperability between mobile devices, Human networks (HUNETs) a network architecture along with B-SUB is used. B-SUB, an interest-driven information sharing system which exploits the peer-to-peer communication pattern in HUNETs, employs content-based networking that achieves infrastructure-less communication between mobile devices. In B-SUB, content and user interests are described by tags, which are human readable strings. Temporal Counting Bloom Filter (TCBF) is invented to encode tags which achieves efficient content routing. Routing in HUNET networking remains an open problem. The main issue is scalability. In this paper, the scalability of the routing scheme, are evaluated by the full Internet AS-level topology and on the internal networks of representative ASes using realistic distributions of content and users extrapolated from traces of popular applications. An experiment is performed to demonstrate the effectiveness of this tag-based content description method and conclude that HUNET is feasible, even with addresses consisting of expressive content descriptors.

1 INTRODUCTION

Traditionally the Internet has been a host- and message-centric system, which has lead to several problems in terms of security, scalability and mobility. Since the sender is in complete control of communication, denial of service attacks are easy to launch. Additionally, an efficient multicast is difficult to implement on the Internet's scale and since the IP address acts as both the node identifier and locator, mobility is problematic to achieve. To overcome these problems, a data-oriented publish/subscribe (pub/sub) networking approach has been proposed.

The notion of content-centric networking is based on an addressing scheme wherein the send and receive communication primitives identify content rather than network locations. The service model of content-centric network supports information pull and push using tag sets as information descriptors. This addressing scheme is motivated by social, application-levelconsiderations, as much as by technical, network-level considerations.

At the network-level, an addressing scheme that identities content as opposed to location would allow the network to operate more efficiently by duplicating and caching content around the network, since it is the delivery of content that matters, not where that content resides. Publish/subscribe is a scalable and flexible communication paradigm which suits the needs of modern applications.

A publish/subscribe service conveys published notifications from any producer to all interested consumers with a matching subscription set. In this manner clients do not use source/destination identifiers or addresses. This inherent loose coupling of producers and consumers is the primary advantage of these systems. A typical pub/sub system consists of publishers, subscribers, and brokers. Publishers, which act as information providers, publish events to brokers. Subscribers, which act as information consumers, express their interests on events by issuing subscriptions to brokers. As service nodes in the network, the functions of brokers are to store, deliver and match of subscriptions and events.

The main purpose of Bloom filters is to build a space efficient data structure for set membership. Indeed, to maximize space efficiency, correctness is sacrificed: if a given key is not in the set, then a Bloom filter may give the wrong answer (this is called a false positive), but the probability of such a wrong answer can be made small. Human networks (HUNETs), a **new a**pplication platform working on top of the networks, formed solely by human-carried wireless devices. An real-world example is Bluejacking [3], where people use bluetooth devices to exchange messages when they are in vicinity.

HUNETs exploits B-SUB (bloom-filter-based publish/SUBscribe), an interest-driven information sharing system. B-SUB exploits the peer-to-peer communication pattern in HUNETs, that routes the messages based on their content of the message, rather than by an explicitly specified destination, without an end-to-end addressing mechanism. An extension to the counting Bloom filter is Temporal Counting Bloom Filter (TCBF), which achieves content routing. B-SUB uses TCBF to encode the tags by compressing content and user's interest and embed information needed for brokers to make forwarding decisions. TCBF are used to store the user interests and its serve a matching hint.

AS-level topology is evaluated to address the routing problem in HUNETs. Usually an AS store all the shortest-paths rooted at every node. The number of users inside each AS depends on the distribution of the users over the AS-level topology. Its supports forwarding along multiple loop-free paths, aggregates addresses for scalability, does not require per-packet network state, and leads to near-optimal paths on average. AS-level topology estimate that incorporates multiple data sources and multiple time periods. The scalability of routing scheme, is evaluated both in terms of memory and computational complexity.

2 SYSTEM OVERVIEW

2.1 HUNETs

A HUNET is composed of portable devices i.e., human-carried mobile devices, that are equipped with wireless communication interfaces, and relies on peer-to-peer communication to do forwarding. Human networks is a dynamic networks composed of human carried wireless devices. HUNETs requires content-based networking services(style of communication that associates publisher and subscriber pairs based on actual content and interests, rather than letting the publisher nodes specify the subscriber). HUNETs enable information sharing between users in a completely decentralized manner without the aid of an wireless communication infrastructure.

DTNs adopt a "store-carry-and-forward" model, which significantly expands the communication capability of mobile device. DTNs do not support interest-driven communication. The routing is based on the end-to-end model, where the information source is unaware of the users who are interested in the information. Many existing DTN routing protocols require complex offline processing to achieve optimal performance. They only allow mobile devices to communicate with each other through wireless infrastructures, for example, GSM/3G/LTE, and so on.

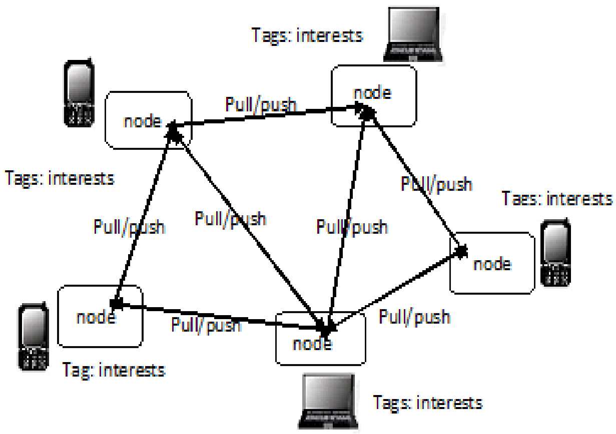

Fig 1: HUNETs Architecture

HUNET, a novel network architecture that facilitates efficient information sharing between portable mobile devices. In PUSH, a node replicates all of the messages in its message buffer to every node that it encounters without duplication. In PULL, a node only forward a message to another node that is interested in the message. B-SUB, an interest-driven information sharing system for HUNETs, a content-based publish/subscribe that achieves infrastructure-less communication between mobile devices. The content-based publish/subscribe limiting the scalability of a network. Driven by the limitations of the publish/subscribe, envision the AS-level topology to address the limiting factor. The service model of a HUNET network supports information pull and push using tag sets as information descriptors. Within this service model, a routing scheme is proposed that supports forwarding along multiple loop-free paths, aggregates addresses for scalability, does not require per-packet network state, and leads to near-optimal paths on average.The TCBF, an extension to the counting Bloom filter to encode tags, which achieves efficient content routing.we present B-SUB-P, an extension of B-SUB that provides stronger privacy guarantee.

2.2 Publish-Subscribe system

In HUNETs architecture, the message pattern where senders sent a message to the receivers with the help Publish-Subscribe system. Here, the senders are called publishers and receivers are called subscribers. In pub-sub, senders of messages do not program the messages to be sent directly to specific receivers. Instead, published messages are characterized into classes, without knowledge of what, if any, subscribers there may be. Similarly, subscribers express interest in one or more classes, and only receive messages that are of interest, without knowledge of what, if any, publishers there are.

In the pub/sub model, subscribers typically receive only a subset of the total messages published. The messages are received based on content-based system. a**content-based** system, messages are only delivered to a subscriber if the attributes or content of those messages match constraints defined by the subscriber. The subscriber is responsible for classifying the messages.

In many pub/sub systems, publishers post messages to an intermediary message broker, and subscribers register subscriptions with that broker, letting the broker perform the process of matching the user interests, subscription and events. The broker normally performs a store and forward function to route messages from publishers to subscribers.

Subscribers may register for specific messages at build time, initialization time or runtime. In GUI systems, subscribers can be coded to handle user commands (e.g., click of a button), which corresponds to build time registration. The frameworks and software products use xml configuration files to register subscribers. These configuration files are read at initialization time. The most sophisticated alternative is when subscribers can be added or removed at runtime.

The Data Distribution Service (DDS) middleware does not use a broker in the middle. Instead, each publisher and subscriber in the pub/sub system shares meta-data about each other via IP multicast. The publisher and the subscribers cache this information locally and route messages based on the discovery of each other in the shared cognizance.

Even with systems that do not rely on brokers, a subscriber might be able to receive data that it is not authorized to receive. An unauthorized publisher may be able to introduce incorrect or damaging messages into the pub/sub system. This is especially true with systems that broadcast or multicast their messages. Encryption can prevent unauthorized access but cannot prevent damaging messages from being introduced by authorized publishers. Architectures other than pub/sub, such as client/server systems are also vulnerable to authorized message senders that behave maliciously. B-SUB-P, an extension of B-SUB that provides stronger privacy guarantee. B-SUB-P removes this openness by altering the basic content routing protocol of B-SUB. B-SUB-P mixes the user’s own interests and relayed interests when two encountered nodes exchange interests.

Pub/sub scales well for small networks with a small number of publishers and subscriber nodes and low message volume. However, as the number of nodes and messages grows, the likelihood of instabilities increases, limiting the maximum scalability of a pub/sub network. Example throughput instabilities at large scales include:

- Load surges—periods when subscriber requests saturate network throughput followed by periods of low message volume (underutilized network bandwidth)
- Slowdowns—more and more applications use the system (even if they are communicating on separate pub/sub channels) the message volume flow to an individual subscriber will slow
- IP broadcast storms—a local area network may be shut down entirely by saturating it with overhead messages that choke out all normal traffic unrelated to the pub/sub traffic.

To overcome these short comings of pub-sub, AS-Level topology along with pub/sub system are been used. AS-level topology estimate that incorporates multiple data sources and multiple time periods. The scalability of the routing scheme is evaluated, both in terms of memory and computational complexity, on the full Internet AS-level topology and on the internal networks of representative ASes using realistic distributions of content and users extrapolated from traces of popular applications.

3 AS-LEVEL TOPOLOGY

Connections between participants in Internet communications can be abstracted in the dimension of network administration, which groups tags into autonomous systems (ASes). An autonomous system that manages one or more networks and has a coherent policy for routing IP traffic both internally and to other ASes.

AS graph are constructed from Internet inter domain TCBF routing tables, which contain the information of connections from an AS to its immediate neighbours. The widely used TCBF data are from the route server which connects to several operational routers within the Internet for the purpose of collecting TCBF routing tables.

Usually an AS store all the shortest-paths rooted at every node. The number of users inside each AS depends on the distribution of the users over the AS-level topology. The growth of the memory requirement is relatively high initially, but steadily attens. This is due directly to the aggregation of tags under routing scheme: even with high numbers of users, the memory required to store all the routing information is likely to remain practically constant, since most of the new descriptors will be aggregated at no additional cost. the AS-level Internet topology is well structured and can be described effectively by the clustering features.

4 SYSTEM DESIGN

4.1 Node Initialization:

The contents of messages and the interests of users are identified by tags, which are strings that summarize the topics of the message. They are stored in TCBFs, which are then used as probabilistic hints for forwarding messages. The pub/sub routing provisions two functions: interest propagation and message forwarding. Both rely on the TCBF to achieve low storage and computational complexities. B-SUB limits the size of messages to a few more than 100 bytes.

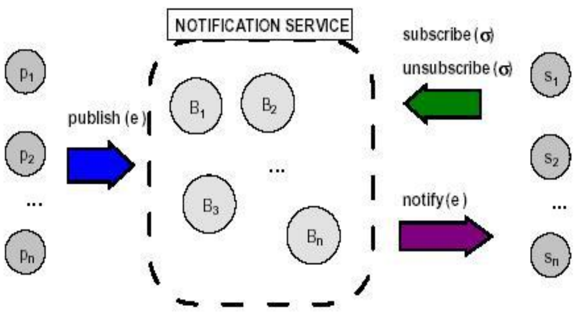

Fig 2: publisher ,broker and subscriber

Three publisher nodes are first created and the distance of each and every node is created. Then a server node, which acts the broker node is then created which contains the TCBF for storing the interests of the server .This server is capable of providing better interest propagation. After this the client node, which acts as the subscriber node is created. The number of user interest depends on the distribution of the interests or the content over the AS-level topology.

4.2 Subscription of the Interested Tag

The subscriber or the client node is capable of viewing the tags of interests. TCBFs are used to compress users' interests. A user stores its own interests in a TCBF .It can subscribe the required tag by sending the query and the broker stores the interests collected from other users in another TCBF. The subscriber identify the publisher of the tag by means of using the search option. Then after giving the submit option, the corresponding publisher node which is the owner of the tag receives the request. TCBFs serve as a "compressed" matching hint for delivery, but they are not precise,

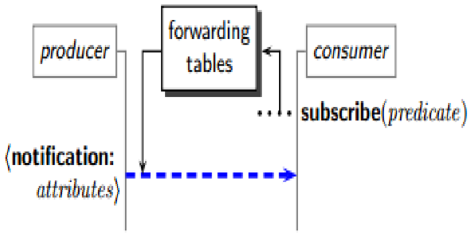

Fig 3: Subscription of interests

4.3 Path Manipulation

AS-level topology quite well that the exact connectivity between the ASes at the level of their routers is not publically available. In this topology it stores all the shortest-paths rooted at every node.

The publisher receives the request from the subscriber with details namely, the subscriber name, the service and also its type it needs, and request status. The routing path between the publisher and the requested subscriber is obtained by using the path manipulation option at the publisher node.

4.3.1 Interest Propagation

In B-SUB, TCBFs are used to compress users' interests. A user stores its own interests in a TCBF, which is called the genuine filter. A broker stores the interests collected from other users in another TCBF called the relay filter. TCBFs serve as a "compressed" matching hint for delivery, but they are not precise, due to the TCBF's false positives. The false positives of the TCBF causes interests being falsely matched by a message, so that the message will be forwarded to nodes that are not really interested in it. Again, false positives may occur. Thus, a delivery tree, instead of a path, will be generated. Nonetheless, it is guaranteed that the original path to the subscriber is embedded in the tree.

4.4 Node Authentication

The neighboring node of the publisher has to be authenticated for security reasons. For this reason the encryption process of the ID of the neighbor node is carried out .Only after this process the neighboring node can join the network for the content routing. Then the encrypted Id is sent to the publisher node for authentication. The decryption of the received ID is carried out and now this neighboring node can join the network. A, B, C, D and E are 32-bit words of the state; *F* is a nonlinear function that varies; $\lll_n$ denotes a left bit rotation by *n* places; *n* varies for each operation; W_t is the expanded message word of round t;K_t is the round constant of round t; $\boxplus$denotes addition modulo 2^{32}.

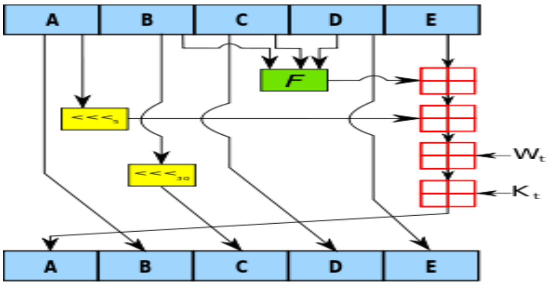

Fig 4: An iteration in SHA-1

4.5 Content Routing With Privacy

The publisher node gives its secret key and from now the TTL of the tag is initialized. The broker or the server node now submits the query to the publisher. The publisher receives the request and it sends the data to the server. Now the server sends the responds to the publisher. After this the publisher sends the content to subscriber or the client by giving its secret key. Finally, the subscriber receives the content of the interested tag.

4.5.1 Privacy guarantee

An extension of B-SUB is used to provides stronger privacy guarantee. The attack model for this discussion is as follows: In our attack model, an attacker disguises as a normal user, and then joins a HUNET and engages in content routing. The attacker collects the interests of other users in doing forwarding. User interests thus are leaked to the attacker and are subjected to privacy infringement.

The basic privacy guarantee provided by the original B-SUB is called nondirect linkage. It means that the attacker cannot obtain direct linkage between a user's identity and his/her interests. This fact could be verified as follows: Given a TCBF that encodes the interests of a user, one cannot discover the real interests of the user because: The attacker cannot reverse back the hashed bit-vector to the real interests. This is provided by the security of the one-way hash functions used in the TCBF.

The attacker can only guess the linkage between the user's identity and his/her interests. B-SUB requires that a user sends a TCBF that encodes its own interests to another user when they encounter. The attacker can guess the real interests by querying the membership of known interests against the TCBF. If an interest is a member of the Bloom filter, then it is assumed to be an interest of the owner of the Bloom filter. This attack achieves the best result when the attacker knows the universal set of interests. Fortunately, this is not true in HUNETs. First, all interests gathered by an attacker are encoded, which means that the attacker cannot directly get the universal set from the users. The attacker can gather other's interests by relaying messages; whenever a message is being delivered to another user, the attacker assumes that the content descriptors of the message are in the universal set. However, this does not establish directly linkage, because the delivery may be caused by a false positive. Without directly examining the raw strings of the user's genuine interests, the attacker cannot associate the interests with users.

5 RESULTS

User interests and message tag. A tag pool of 38 tags is prepared. Every node is assigned two tags as its interests, which are selected uniformly from the tag pool. Every node generates messages constantly. The interval between the generation of two successive messages is drawn evenly from TCBF

Routing messages on the paths that have the minimal delay improve the overall delivery performance. For the large number of content based descriptors, the scalability of the routing scheme are evaluated in terms of both memory requirements and the cost for maintaining the routing information. The fundamental routing problem in HUNETs which are in the information-centric network are studied with the help of AS-level topology.

6 CONCLUSION

In this project the concept of Human Networks have been created by means of creating multiple publishers and multiple subscribers who are connected with each other at a specified distance. Routing in HUNET networking remains an open problem. The main issue is scalability. Traditional IP routing can be used with name prefixes, but it is believed that the number of prefixes will grow too large. A related problem is the use of per-packet in-network state. A routing scheme is proposed to solve these problems. The service model of the HUNET network supports information pull and push using tag sets as information descriptors. Within this service model, a routing scheme is proposed that supports forwarding along multiple loop-free paths, aggregates addresses for scalability, does not require per-packet network state, and leads to near-optimal paths on average. The scalability of routing scheme, is evaluated both in terms of memory and computational complexity. HUNET is feasible, even with addresses consisting of expressive content descriptors.

REFERENCES

[1] B.H. Bloom, "Space/Time Trade-Offs in Hash Coding with Allowable Errors,"Comm. ACM,vol. 13, no. 7, pp. 422-426, 1970.

[2] W. Gao and G. Cao, "User-Centric Data Dissemination in Disruption Tolerant Networks," IEEE INFOCOM, pp. 3119-3127, Apr. 2011.

[3] W. Fang, F. Liu, F. Yang, L. Shu, and S. Nishio, "Energy-Efficient Cooperative Communication for Data Transmission in Wireless Sensor Networks,"Consumer Electronics, IEEE Trans.,vol. 56, no. 4, pp. 2185-2192, Nov. 2010.

[4] G. Sollazzo, M. Musolesi, and C. Mascolo, "TACO-DTN: a Time-Aware Content-Based Dissemination System for Delay Tolerant Networks,"Proc. ACM First Int'l MobiSys Workshop Mobile Opportunistic Networking (MobiOpp '07),pp. 83-90, 2007

[5] K. Fall, "A Delay-Tolerant Network Architecture for Challenged Internets," Proc. Conf. Applications, Technologies, Architectures, and Protocols for Computer Comm., pp. 27-34, 2003.

[6] Y. Li, G. Su, P. Hui, D. Jin, L. Su, and L. Zeng, "Multiple Mobile Data Offloading through Delay Tolerant Networks," Proc. Sixth ACM Workshop Challenged Networks, pp. 43-48, http://doi. acm.org/10.1145/2030652.2030665, 2011.

[7] Y. Zhao and J. Wu, "B-Sub: A Practical Bloom-Filter-Based Publish-Subscribe System for Human Networks,"Proc. Int'l Conf. Distributed Computing Systems (ICDCS '10),pp. 634-643, 2010.

[8] Yaxiong Zhao, and Jie Wu, "The Design and Evaluation of An Information Sharing System for Human Networks"IEEE transactions on parallel and distributed systems, vol. 25, no.3, march 2014.

[9] Michele Papalini, Antonio Carzaniga, Koorosh Khazaei and Alexander L. Wolf, "Scalable Routing for Tag-Based Information-Centric Networking"

[10] Yan Li, Jun-Hong Cui, Dario Maggiorini and Michalis Faloutsos, "Characterizing and Modelling Clustering Features in AS-Level Internet Topology"

[11] Shi Zhou & Raul J. Mondragon, " Analyzing and modelling the AS-level Internet topology"

[12] T. Pongthawornkamol, K. Nahrstedt, and G. Wang, "The Analysis of Publish/Subscribe Systems over Mobile Wireless Ad Hoc Networks," pp. 1-8, Aug. 2007.