OMiN - Opportunistic Microblogging Network

Neil Wells, supervised by Tristan Henderson

Abstract

OMiN is a pocket switched network running on smartphones. It allows users to send and receive microblog messages without using any global infrastructure such as the internet. Smartphones in close proximity to each other pass on messages according to a variation of the PROPHET routing protocol. Steps have been taken to verify the security of the network by protecting it against known attack vectors.

Declaration

I declare that the material submitted for assessment is my own work except where credit is explicitly given to others by citation or acknowledgement. This work was performed during the current academic year except where otherwise stated.

The main text of this project report is NN,NNN* words long, including project specification and plan.

In submitting this project report to the University of St Andrews, I give permission for it to be made available for use in accordance with the regulations of the University Library. I also give permission for the title and abstract to be published and for copies of the report to be made and supplied at cost to any bona fide library or research worker, and to be made available on the World Wide Web. I retain the copyright in this work.

Contents

Introduction

TODO

Objectives

Primary Objectives

- design and implement a protocol for discovering nodes in close proximity and passing messages and necessary metadata between them.
- Create a core library to manage message storage and routing.
- Implement a simple epidemic routing algorithm to send messages to all available nodes.
- Design a routing algorithm using user metadata to route messages while disguising message content and metadata.

Secondary Objectives

- Create a smartphone UI.
- Implement a more advanced routing algorithm.
- Design and implement a mechanism to decide whether a node is trustworthy or not.
- Evaluate the performance of the implemented routing algorithms.

Tertiary Objectives

- Compare the real world vs simulated performance of the routing algorithms.

Context Survey

The following provides brief summary of the current state-of-the-art in opportunistic network technology. Only the most relevant subjects will be addressed in order to give the reader sufficient background information to fully understand the project.

Opportunistic Networks

An opportunistic network is a network where connections between nodes are sparse and a direct path from source to destination cannot be guaranteed. For example, a common form of opportunistic network (and the form we will focus on) is the Pocket Switched Network (PSN) - a network of smartphones carried around by people. Because of the predictable nature of human behaviour, much research has been done to improve PSN algorithms.

Opportunistic networks must be able to store messages and forward them when connections become available. Problems that have been solved in conventional connected networks (security, routing etc.) are much harder to solve in opportunistic networks.

Similar Projects

Haggle

One of the largest opportunistic platforms is Haggle (http://www.haggleproject.org)[1], a pocket switched network designed to run on smartphones. There are implementations for a number of clients including Android (play.google.com/store/apps/details?id=org.haggle.kernel) and Windows Mobile.

By monitoring use of the platform, the authors discovered trends in inter-contact times and contact durations, showing that existing algorithms are poorly suited to real world models[2].

FireChat

FireChat (opengarden.com/firechat) is a new smartphone application used for off-the-grid messaging between nearby users. It has been used to circumvent government restrictions in Iraq (http://www.theguardian.com/technology/2014/jun/24/firechat-updates-as-40000-iraqis-download-mesh-chat-app-to-get-online-in-censored-baghdad) and during the Hong Kong protests (http://www.theguardian.com/world/2014/sep/29/firechat-messaging-app-powering-hong-kong-protests).

However, the app mostly relies on an internet connection, and its simple protocol is insecure (http://breizh-entropy.org/~nameless/random/posts/firechat_and_nearby_communication) and unable to implement the store-and-forward functionality of a proper opportunistic network.

SWIM

The Shared Wireless Infostation Model (SWIM) proposes using an opportunistic network to monitor whales[3]. Small nodes are attached to the whales, which record data such as location and interaction with other whales. Connected nodes transfer this data between each other. Whenever data is transferred to a base station (the paper proposes using seabirds), it can be collected and stored.

Because data is shared between nodes, it is no longer necessary to find a whale with a sensor in order to acquire data from that sensor. This is a perfect example of the power of opportunistic networks in an environment with very limited connectivity.

Routing Algorithms

Routing in opportunistic networks can be reduced to the problem of finding a destination node in a constantly changing graph. Most routing algorithms therefore build on existing graph search techniques. However, because the graph is constantly changing and is not necessarily random, such techniques are not necessarily the most effective (as shown by the Haggle project).

Context Based Routing

Context based routing is a form of greedy searching, where a single message is continually passed to the node most likely to reach the destination. There are a variety of methods to compute the utility of a node, including CAR[12] and MobySpace[13]. While it is not guaranteed to find the optimum path to the destination, it uses very few resources as there is only one message being stored.

Epidemic Routing

The opposite of context based routing is epidemic routing - a form of uniform cost search[4]. Copies of the message are passed at every opportunity until the network is saturated. This is often likened to the spread of a virus. While this approach will always find the optimal path (because it takes all possible paths), it is very resource intensive - all nodes are expected to store every possible message. For this reason, routing protocols that use similar techniques (dissemination based routing) concentrate on avoiding unnecessary use of resources.

PROPHET

The Probabilistic Routing in Intermittently Connected Networks (PROPHET) algorithm[5] uses a form of the A* algorithm. A utility function (derived from recent encounters with nodes) is used to predict whether the message should be passed on

to a node. This use of heuristics greatly improves performance when compared to epidemic routing.

Bubble RAP

The Haggle project discovered that algorithms that treat routing as a generic graph search problem are often unsuited to PSNs. Bubble RAP[6] works on the idea that a social connections graph has tree like structure, where close nodes form a community. In order to send messages to a different community, the message is moved up the tree towards highly connected nodes near the root, and then back down the tree towards the destination community and, eventually, the destination node.

This has been shown using the data collected from Haggle to be more effective than standard routing algorithms.

Security

Opportunistic networks (especially pocket switched networks) must consider the possibility of an unsecure node which does not conform to the protocol specification (either through accident or design) or attempts to discover private information. Common attack types include:

- Sybil attacks: impersonating another node in order to send messages that appear to be from that node or to receive messages intended for the node.
- Majority attack: by controlling a large number of nodes, an attacker can control a network which assumes that the majority of nodes can be trusted.
- Eavesdropping: gathering information such as message metadata to discover private information such as message contents and user location.
- Denial of Service: saturating the network with unwanted messages.
- Blackhole attack: failing to pass on messages to either reduce resource usage or as part of another attack.

Trust Based Security

Trust based security mechanisms depend on generating a list of trusted or untrusted nodes. This is commonly based on trusting 'friends' in a social network[8] or distrusting nodes exhibiting strange behaviour[9]. While trust based security mechanisms can lower the chances of an attack, they are not infallible.

Certificate Based Security

In order to increase security, a security mechanism must reduce the number of trusted parties to an absolute minimum. This can be done using an infrastructure of public key certificates to verify the identity of a node. However, the problem of distributing this infrastructure without trusting an arbitrary number of nodes has not been solved yet. On the internet, it is done by trusted certificate authorities, who manage and release certificates. However, having a central trusted authority is infeasible in an opportunistic network, as all nodes would have to connect to it directly at some point. Some mechanisms, like the one proposed by Shikfa et al[10] do use a central server, but only require it to be available for nodes joining the network. Other mechanisms split the sdf. Mechanisms for distributed certificate distribution require some level of trust in network nodes. For example Capkun et al's approach[11] does this by building a graph of certificates determining who trusts who. Any abnormalities in the trust graph may indicate foul play. However, this scheme is still vulnerable to a majority attack.

Identity Based Security

Identity based encryption is an increasingly common form of encryption where the user's details (such as a username) acts as a public key, and a private key will be generated by a central private key generator (PKG). However, having a central PKG is problematic in a distributed network. Some security frameworks assume that there is a central PKG that can and will be accessed occasionally[14]. Others split up the PKG into multiple nodes, all of whom must collaborate to generate a private key[15] or split users into communities, each with their own PKGs[16].

Requirements Specification

User Requirements

Non-Functional Requirements
- High: The user shall be able to create a unique identity.
- High: The user shall be able to send text messages to all others who follow the user or a hashtag in the message.
- High: The user shall be able to 'follow' any user and receive messages sent by that user.
- Low: The user shall be able to 'follow' any hashtag and receive messages containing that hashtag.
- Low: The user shall be able to send encrypted direct messages to a single user.

- Low: The user shall be able to send multimedia messages to all others who follow the user or a hashtag in the message.

Functional Requirements
None

System Requirements

Non-Functional Requirements

- High: The system shall work on smartphones or tablets capable of connecting to a wifi network.
- High: The system shall allow creation of user identities with a unique cryptographic identity.
- High: The system shall automatically connect to nearby nodes and pass on relevant information such as messages.
- Medium: The system shall provide a mechanism for securely distributing the cryptographic identity of a user.
- Medium: The system shall protect user metadata such as location and friends list from all other nodes.
- Medium: The system shall ensure that messages cannot be modified in transit or that such modifications can be detected.
- Medium: The system shall ensure that nodes cannot send a message that appears to be from another user.
- Medium: The system shall be robust and able to continue functioning when it encounters an unexpected state such as a malfunctioning or untrustworthy node.
- Medium: The system shall ensure that encrypted direct messages cannot be read by third parties.
- Low: The system shall be able to support multiple user identities on a single node.

Functional Requirements

- High: The system shall collect anonymous logging data for debugging and profiling purposes.
- Medium: The system shall ensure that a reasonable number of messages reach their intended recipients.
- Medium: The system shall restrict the size of the message buffer by evicting messages.
- Medium: The system shall use a minimal amount of the available power.

Software Engineering Process

TODO

Ethics

In order to test the real world performance of the network, we may ask people to use the application. In this case, some metadata will be collected on users, with their consent. This may include:

- An anonymous user ID.
- Anonymised 'Friends list' (or equivalent) of users.
- Times and locations of encounters between anonymous users.
- Metadata of messages passed during encounters, including message ID and origin ID, but NOT message contents.

Design

Use cases

We propose the following use cases for our network:

Disaster Area

A tsunami has wiped out all communications infrastructure in the area and injured a lot of people. Our opportunistic network is the quickest way contact medical teams to inform them of injured people who need help. We can assume that most people have smartphones and will be moving about regularly. Alice is injured and must contact the nearest free medical team so that they can help her. She uses her smartphone to publish a message with the#medical hashtag, her location and status to all nearby nodes. The message is distributed in this manner until it reaches doctor Bob, who subscribes to #medical. Bob sends a reply message to indicate that help is coming and goes to help Alice. Once she is happy that the relevant help is coming, she sends out another message to indicate that she no longer needs help. Doctor Carol, who was also on her way after receiving the first message now knows that she is no longer required.

Privacy

A group of activists are concerned that their internet access is being monitored by their government and their online identity could be hacked or blocked in order to oppress them. They use our network to communicate in a way that cannot be blocked or subverted while revealing as little metadata (such as location) as possible. In order to organise a protest, Alex, a well known activist whose identity is unknown, sends out a message with details of the protest. Brian cannot come at that time because he has an exam, so he informs Alex of this via a secure direct message. Alex sends out a new message with a new date for the protest.

Threats

From the use cases, we can construct a model of potential threats and their motives.

Disaster Area

In a disaster area, people tend to act altruistically. Therefore they are unlikely to attempt to subvert the network. However, people may act selfishly by attempt to conserve battery or memory space on their phone.

- Goal: Selfishly reduce personal resource usage
  - Avoid accepting messages to be passed on (blackhole attack)
  - Avoid making connections unless it directly benefits the user (e.g. to allow the user to send a message)

Privacy

In this scenario, it is very important that the network cannot be compromised by attackers. These attackers may have significant resources available to them.

- Goal: Prevent a message from being disseminated
  - Force all nodes to discard message
    - Control most network nodes (majority attack)
      - Create lots of blackhole nodes
      - Gain control of nodes
    - Overload most nodes (denial of service attack)
- Goal: Modify message while in transit
  - Control a node in the message path
    - Create node and get it into the message path
    - Gain control of a node in the path
- Goal: Prevent a user from sending messages

- ○ Force all nodes to discard messages from user
    - ■ Control most network nodes (majority attack)
        - ● Create lots of blackhole nodes
        - ● Gain control of nodes
    - ■ Discredit user
        - ● Send fake messages from user (sybil attack)
        - ● Exploit trust mechanism
- ○ Hack user's device
- ○ Legal action
    - ■ Discover user identity
- ● Goal: Identify physical identity of a user
    - ○ Derive network topology and route taken by a message to identify user location
    - ○ Identify unique information about a user (such as phone number)
- ● Goal: Identify contents and recipient of an encrypted message
    - ○ Derive route taken by a message to identify recipient
    - ○ Discover private encryption key of recipient
        - ■ Control master PKG
        - ■ Brute force attack

Authentication Models

Authentication is a major issue in opportunistic networks: how can we be sure that a message was sent by a user, and has not been modified.

The simplest option is to forfeit authentication entirely and trust every node to behave correctly. This is possible for small networks, but as the network grows, the chances of having a deceitful user grows (even if each user is 99.99% trustworthy, a network with 7000 users has a 50% probability of having a deceitful user). In order to build a scalable network, we must discount this strategy as being unfeasible.

Networks which may contain deceitful users must use measures to prevent this. A common approach is for each user to identify nodes which may be trustworthy or untrustworthy. Social authentication is a common strategy: only 'friends' of users are considered trustworthy. The IRONMAN algorithm takes a different approach by detecting nodes displaying unusual behaviour. These nodes are then marked as untrustworthy. This approach of using heuristics to separate trustworthy and untrustworthy nodes is useful, but still susceptible to attack.

The best approach to avoid trusting an attacker is to trust as few nodes as possible. The only way to verify that a message actually came from a user is to have some certificate irrefutably proving that a message was sent by a user. If a message comes with a cryptographic certificate then anyone with the sender's public key can verify its origin. So we now encounter a new problem: how can we distribute a public key while trusting as few nodes as possible?

Key Distribution

In order for a key based approach to work, we must be able to verify that an arbitrary user identifier (i.e. a username chosen by the user) is uniquely associated with an arbitrary (randomly chosen) public key. Modern Internet based techniques involve a trusted certificate authority (CA) who will distribute certificates guaranteeing that a key is associated with a user. This is impractical in a pure opportunistic network setting (totally free of infrastructure), as it involves every node having contact with the CA, which considerably restricts the size of the network.

Instead of formally verifying identities, we could build up a trust based infrastructure for distribution of public keys. While this is a common method used in opportunistic networks, trust based approaches cannot be 100% secure (in particular, they are vulnerable to attackers controlling the majority of nodes).

In cases where a sporadic internet connection is available, a central CA can be used to allow identity based encryption by generating a private key for a user, so their identity (username) can be used as their public key. This has the added advantage of preventing sybil attacks because the CA can ensure that every user has a unique username.

Every node with a private key is capable of becoming a PKG and issuing private keys to other users. If the master PKG isn't available via the internet, another node can act as a delegate PKG. In this way we can create a chain of key generators where the master PKG (accessible via the internet) delegates PKG responsibilities down the chain. A node's private key will be compromised if one of its parents is compromised, so it is wise to keep this chain as short as possible. The risk of being compromised can be reduced by allowing users to get private keys from multiple PKGs. If all private keys are needed to verify a user, this reduces the risk of a node n being compromised by its PKG parents.

We can quantify the risk of a node being compromised with the formula

$C(n) = 1 - (1 - V)(1 - \prod_{p \in P(n)} C(p))$ where the chance of a node being untrustworthy is V,

$C(master) = V$, and $P(n)$ is the set of parent PKGs of n. As long as $0 < V < 1$, we believe (although it has not been proven) that the exact value of V will not affect the relative trustworthiness of two nodes (i.e. if $\exists\, 0 < V < 1$ such that $C(a) > C(b)$ then $C(a) > C(b)\ \forall\ 0 < V < 1$). Nodes should always seek to decrease this risk by finding new PKGs.

This scheme has the following advantages:

- It is scalable to any number of users
- Users with the same identity can coexist iff they avoid using the same PKGs
- The probability of a user being compromised can be directly quantified
- If all users can connect to the master PKG over the internet, the scheme simplifies to normal (totally secure) ID based certification

Blackhole Attacks

TODO

Routing Algorithm

Now we come to the problem of routing. Simply put, a routing algorithm takes all available information about a message and uses that information to decide where to send it. In trust based networks, this information is freely available to all trusted nodes (friends lists, recipients, previous paths etc). However, we have opted for a model where all nodes are considered untrustworthy. This means that we must obscure or remove all of this information, while still allowing a routing algorithm to use it.

Our network will be sending messages to multiple users, so a dissemination based routing protocol is advised. Given the disadvantages of epidemic routing, we will use a variation of the PROPHET routing algorithm[5] to distribute messages. We can use a variation of the SSNR-OSNR algorithm[17] to obfuscate sensitive metadata.

Message Buffer Eviction

When the message buffer is too large, it must evict a message. This message should be the least likely to be used again. The simplest approach is the LRU (Least Recently Used) algorithm. Here, the oldest messages are removed on the assumption that they are probably the most widely distributed.

However, we have better ways of measuring message distribution. Give each message a distribution score of how many times the message has been sent or a node is encountered which already has the message. At eviction time, the message with the highest distribution count is evicted.

Implementation

TODO

Evaluation and Critical Appraisal

TODO

Conclusions

TODO

References

[1] Scott J, Crowcroft J, Hui P, Diot C, Others. Haggle: A networking architecture designed around mobile users. In: WONS 2006: Third Annual Conference on Wireless On-demand Network Systems and Services; 2006. p. 78-86.

[2] Chaintreau A, Hui P, Crowcroft J, Diot C, Gass R, Scott J. Impact of human mobility on opportunistic forwarding algorithms. Mobile Computing, IEEE Transactions on. 2007;6(6):606-620.

[3] Small T, Haas ZJ. The Shared Wireless Infostation Model: A New Ad Hoc Networking Paradigm (or Where There is a Whale, There is a Way). In: Proceedings of the 4th ACM International Symposium on Mobile Ad Hoc Networking &Amp; Computing. MobiHoc '03. New York, NY, USA: ACM; 2003. p. 233-244. Available from: http://doi.acm.org/10.1145/778415.778443.

[4] Vahdat A, Becker D, Others. Epidemic routing for partially connected ad hoc networks. Technical Report CS-200006, Duke University; 2000.

[5] Lindgren A, Doria A, Schelén O. Probabilistic Routing in Intermittently Connected Networks. SIGMOBILE Mob Comput Commun Rev. 2003 Jul;7(3):19-20. Available from: http://doi.acm.org/10.1145/961268.961272.

[6] Hui P, Crowcroft J, Yoneki E. Bubble Rap: Social-based Forwarding in Delay Tolerant Networks. In: Proceedings of the 9th ACM International Symposium on Mobile Ad Hoc Networking and Computing. MobiHoc '08. New York, NY, USA: ACM; 2008. p. 241-250. Available from:http://doi.acm.org/10.1145/1374618.1374652.

[7] Pelusi L, Passarella A, Conti M. Opportunistic networking: data forwarding in disconnected mobile ad hoc networks. Communications Magazine, IEEE. 2006 Nov;44(11):134-141. Available from: http://dx.doi.org/10.1109/MCOM.2006.248176.

[8] Trifunovic S, Legendre F, Anastasiades C. Social Trust in Opportunistic Networks. In: INFOCOM IEEE Conference on Computer Communications Workshops, 2010; 2010. p. 1-6. Available from: http://dx.doi.org/10.1109/INFCOMW.2010.5466696.

[9] Bigwood G, Henderson T. IRONMAN: Using Social Networks to Add Incentives and Reputation to Opportunistic Networks. In: Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on; 2011. p. 65-72. Available from: http://dx.doi.org/10.1109/PASSAT/SocialCom.2011.60.

[10] Shikfa A, Onen M, Molva R. Bootstrapping security associations in opportunistic networks. In: Pervasive Computing and Communications Workshops (PERCOM Workshops), 2010 8th IEEE International Conference on; 2010. p. 147-152. Available from:http://dx.doi.org/10.1109/PERCOMW.2010.5470676.

[11] Capkun S, Buttyan L, Hubaux JP. Self-organized public-key management for mobile ad hoc networks. Mobile Computing, IEEE Transactions on. 2003 Jan;2(1):52-64. Available from: http://dx.doi.org/10.1109/TMC.2003.1195151.

[12] Musolesi M, Hailes S, Mascolo C. Adaptive routing for intermittently connected mobile ad hoc networks. In: World of Wireless Mobile and Multimedia Networks, 2005. WoWMoM 2005. Sixth IEEE International Symposium on a; 2005. p. 183-189. Available from:http://dx.doi.org/10.1109/WOWMOM.2005.17.

[13] Leguay J, Friedman T, Conan V. Evaluating Mobility Pattern Space Routing for DTNs. In: INFOCOM 2006. 25th IEEE International Conference on Computer Communications. Proceedings; 2006. p. 1-10. Available from: http://dx.doi.org/10.1109/INFOCOM.2006.299.

[14] Kamat P, Baliga A, Trappe W. An Identity-based Security Framework For VANETs. In: Proceedings of the 3rd International Workshop on Vehicular Ad Hoc Networks. VANET '06. New York, NY, USA: ACM; 2006. p. 94-95. Available from: http://doi.acm.org/10.1145/1161064.1161083.

[15] Kong J, Petros Z, Luo H, Lu S, Zhang L. Providing robust and ubiquitous security support for mobile ad-hoc networks. In: Network Protocols, 2001. Ninth International Conference on; 2001. p. 251-260. Available from:http://dx.doi.org/10.1109/ICNP.2001.992905.

[16] Tseng FK, Zao JK, Liu YH, Kuo FP. Halo: A Hierarchical Identity-Based Public Key Infrastructure for Peer-to-Peer Opportunistic Collaboration. In: Mobile Data Management: Systems, Services and Middleware,

2009. MDM '09. Tenth International Conference on; 2009. p. 672-679. Available from: http://dx.doi.org/10.1109/MDM.2009.115.

[17] Parris I, Henderson T. Privacy-enhanced social-network routing. Computer Communications. 2012;35(1):62-74. Available from:http://www.sciencedirect.com/science/article/pii/S0140366410004767.

Appendices

Appendix A - Testing Summary

TODO

Appendix B - Status Report

TODO

Appendix C - User Manual

TODO

Appendix D - Maintenance Document

TODO