Random Key Predistribution Schemes for Sensor Networks

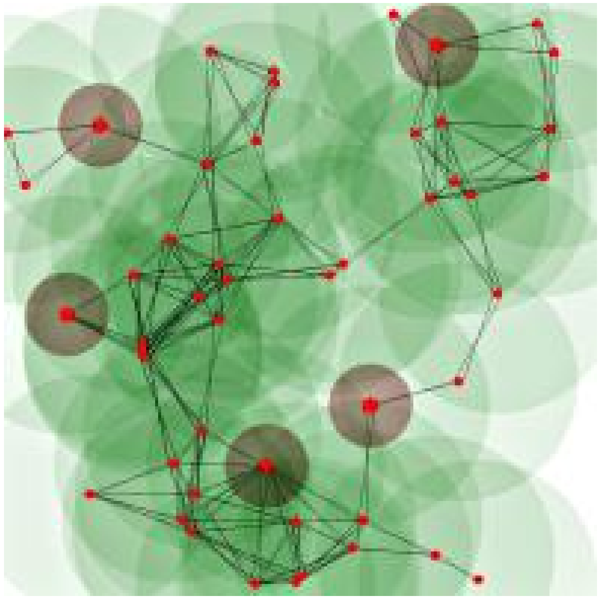
Haowen Chan, Adrian Perrig, Dawn Song
Carnegie Mellon University

Problem Domain

- Boot strapping protocol
  - Secure infrastructure for newly deployed sensor network
  - Network discovery by newly deployed nodes
- Mechanisms to provide a secure infrastructure for newly deployed nodes
  - Q-composite random key pre distribution
  - Multi-path key reinforcement
  - Random pairwise keys

Network Architecture

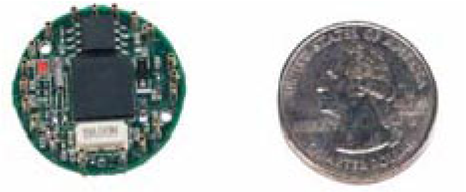
- Physical installation or random scattering
- Sensor network is deployed by a single party
- Communication
  - Node 2 node, node 2 base station, base station to node
- Deployment density
  - Network size > 1000 nodes
  - 20+ neighbours within communication range



- *Deployed by a single independent party*

Revisiting the challenges

- Public key cryptosystems are expensive
- Vulnerability to physical capture
- No prior knowledge of post-deployment configuration
- Limited
  - Memory, bandwidth, transmission power
- Over reliance on base stations
  - Fast Response
  - Limited Flexibility



http://www.zess.uni-siegen.de/cms/front_content.php?idcat=76

Evaluation Metrics

- Resilience against node capture
  - Fraction of total network communications compromised
- Resilience against node replication
- Revocation
- Scale

Basic Random Key Pre-distribution

- Proposed by Eschenauer and Gligor
- Initialisation Phase: Setting up the Key Ring for each node
  - Select S random keys from the total possible key space
  - Randomly select m keys from S for each node
- Key-setup
  - Key discovery through broadcast of key identifiers
  - Neighbour verification through challenge-response protocol(s)
  - Use of the shared key as the link key
  - Path key set-up
    - Between nodes in the 'vicinity' that do not share a common key

Probabilistic modelling

- Expected degree (d) such that a graph is 'connected' with a high probability (c)

$$d=(n-\frac{1}{n})(\ln(n)-\ln(-\ln(c)))$$

- Probability (p) of successfully performing a key set-up with some neighbour and the expected number of neighbours in communication range (n')

$$p=d/n'$$

- Range extension
  - Detection of connectivity at node
  - Increase transmission power
  - Request neighbours to forward communication for a few hops

Q Composite Keys Scheme

Q Shared keys instead of 1

- Initialisation Phase: Setting up the Key Ring for each node
  - Select S random keys from the total possible key space
  - Randomly select m keys from S for each node
- Key-setup
  - Discover all common keys between node and neighbour
    - A single broadcast of key identifiers
    - Use of Merkle puzzles (susceptible to man in the middle attacks)
  - Identify neighbours with more than q keys in common
  - Communication link key $K = hash(k_1 \| K_2 \| \ldots \| k_{q^1})$
    - Keys are hashed in some (predetermined) canonical order

Computation of Key Pool Size

- Known parameters
  - Network size (n)
  - Probability of full network connectivity (c)
  - Expected number of neighbours in communication range ($n^{|}$)
  - No of keys in Key Ring (m)
- Calculate
  - Expected degree of each node (d)
  - Desired probability that any two nodes can perform key set-up (p)
- Calculate S, probability of two nodes sharing at least q keys ≥ p
- *Trade-off*: Higher probability of Key establishment *vs.* Security

Computation of Key Pool Size (2)

- Let p(i) be the probability that two nodes have i keys in common
- Total no of ways for a node to pick m keys = $\binom{|S|}{m}$
- Total no of ways for both nodes to choose m keys = $\binom{|S|}{m}^2$
- Total no. of ways to choose i common keys = $\binom{|S|}{i}$
- Total no of ways to choose the remaining 2(m-i) disjoint keys = $\binom{|S|-i}{2(m-i)}$
- No of ways of partitioning the disjoint keys = $\binom{2(m-i)}{m-i}$

Computation of Key Pool Size (3)

- Finally, we have:

$$p(i)=\frac{\left(\binom{|S|}{i}\binom{|S|-i}{2(m-i)}\binom{2(m-i)}{m-i}\right)}{\binom{|S|}{m^2}}$$

- Let p_{connect} be the probability of two nodes sharing sufficient keys to form a secure connection.
- $p_{connect}=1-(p(0)+p(1)+\cdots+p(q-1))$
- Given m, q, p we need to maximise |S| such that $p_{\text{connect}} \geq$ p

Node Revocation

- *From Eschenauer and Gligor:* Some salient features
  - Ability to revoke the entire key ring of a compromised node.
  - A controller node broadcasts a single revocation message containing a signed list of k key identifiers for the key ring to be revoked.
  - The controller generates a signature key K_e and unicasts it to each node by encrypting it with a key K_{ci}
  - Re-Keying: in the rare case that the lifetime of a key expires
    - self-revocation of a key by a node.
    - No network-wide broadcast
    - affected nodes restart the shared-key discovery and, possibly, the path-key establishment, phase.

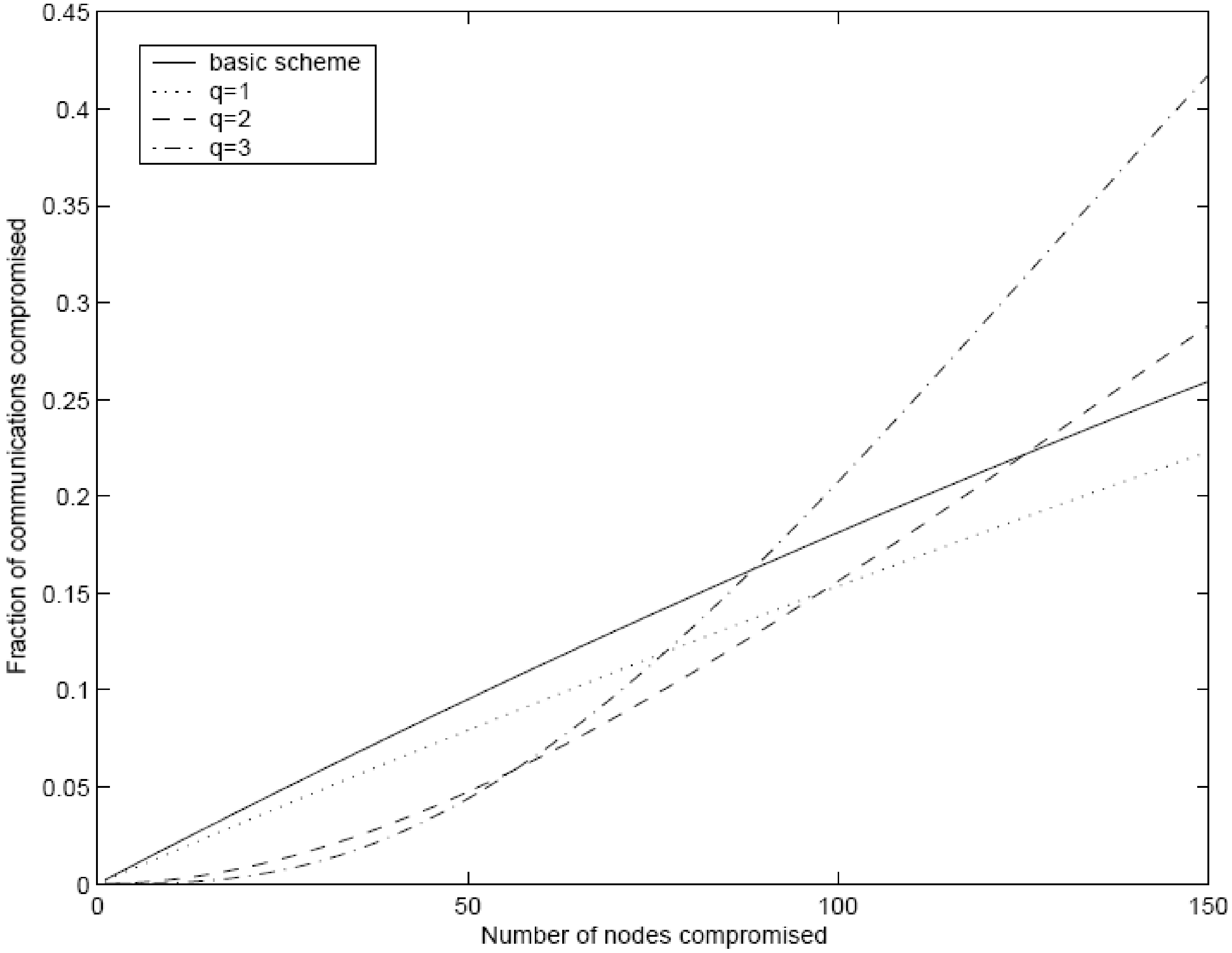
Resilience against node capture

- Fraction of network links compromised as a result of note capture.
- Let no of nodes captured be x, each node with m keys
- Possibility that a given key has not been compromised = $\left(1-\frac{m}{|S|}\right)^x$
- For a communication link whose link key is the hash

of *i* shared keys, probability of link being compromised = $\left(1-\left(1-\frac{m}{|S|}\right)^x\right)^i$

- Probability of setting up a secure link $p = p(q) + p(q+1) + \cdots + p(m)$
- Therefore, probability that any secure link is compromised is $\sum_{i=q}^{m}\left(1-\left(1-\frac{m}{|S|}\right)^x\right)^i \frac{p(i)}{p}$

Evaluation ... contd



- Not infinitely scalable
- Greater resilience only when no of captured nodes is small
- Removes incentive for small scale attacks

However, no resistance against node replication

Maximum Supportable network sizes

- Compromise Threshold – f_m
- Limited global payoff requirement
  - Every subsequent capture reveals no more than communication links than average connectivity degree of a single node
- For x compromised nodes, some fraction f(x) of the communications links are compromised
- Let x_m be the number of compromised nodes such that $f_m = f(x_m)$
- The adversary holds $x_m d$ connections
- Additional links compromised must be less than $x_m d$

$$\left(\frac{nd}{2} - x_m d\right) f(m) \leqslant x_m d$$

Maximum Supportable network sizes

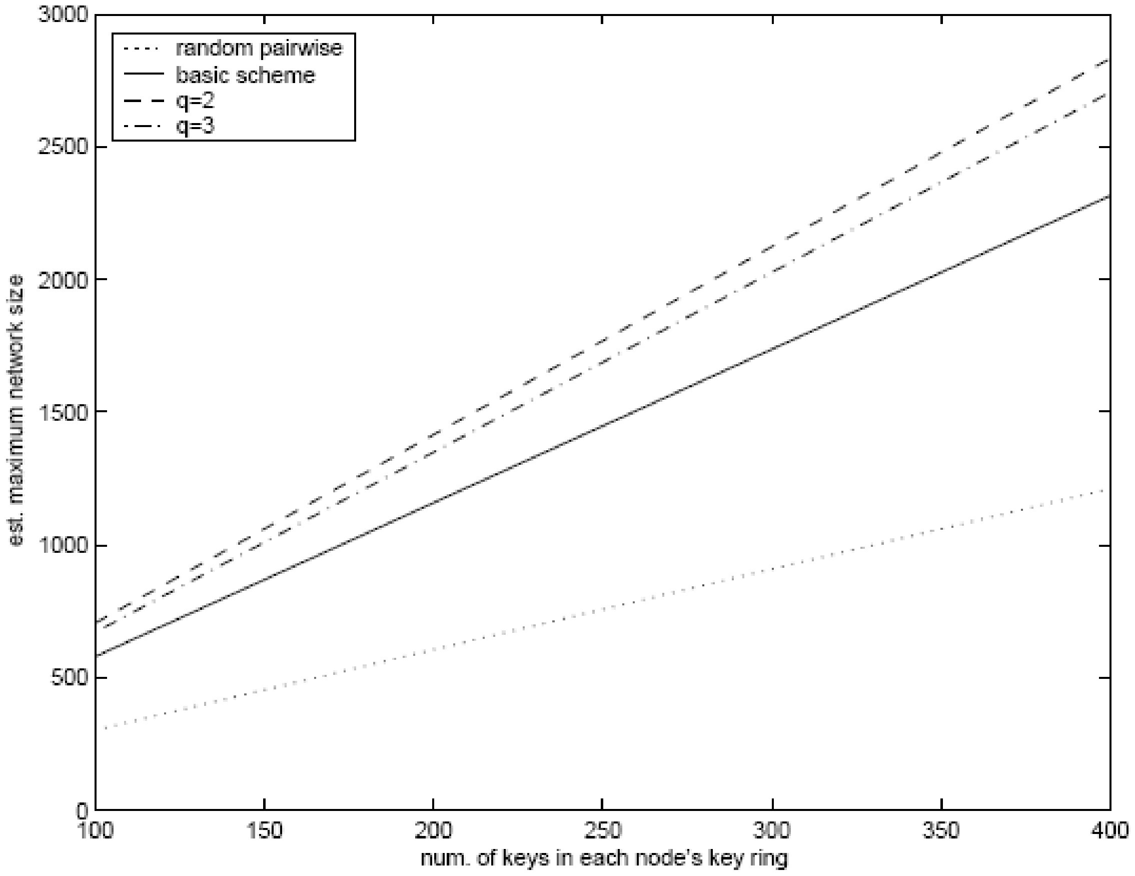


Figure 3. Maximum network sizes
$(p = 0.33, f_m = 0.1)$

$$\left(\frac{nd}{2} - x_m d\right) f(m) \leqslant x_m d$$

Simplifies to

$$n \leqslant 2\mathrm{x}_m \left(1 + \frac{1}{f_m}\right)$$

Multipath Key Reinforcement

- Secure paths set-up using basic random key scheme
- Update key value to a random value
- Utilise multiple independent paths in addition to direct link
  - Source A must be aware of some disjoint paths to B (≤ k hops)

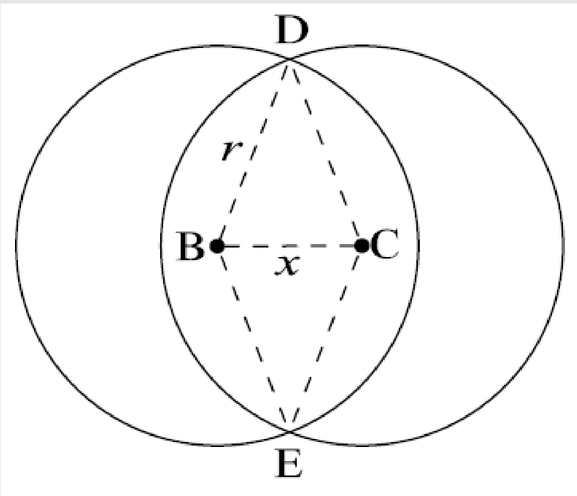
    $k' =$ □
  - Enough routing information must be exchanged during key set-up
- A generates random values via the set of disjoint paths to B
- New link key $k^1 = k \oplus v_1 \oplus v_2 \oplus \ldots \oplus v_j$

Multipath Key Reinforcement (2)

- More paths – increased security
- Length of path
  - Probability of an eavesdropper
  - Weakest link
  - Communication overhead
- 2-hop multipath key reinforcement
  - Minimum path discover overhead
  - Simply look for common neighbours

Effectiveness

- A first step is to calculate the number of expected neighbours



$$A(x) = 2r^2 \cos^{-1}\left(\frac{x}{2r}\right) - x\sqrt{r^2 - \frac{x^2}{4}}$$

If d=20, n1 = 60 then, Expected no of reinforcing neighbours = 3.83

- Probability distribution function of distance between two nodes in communication radius

$$F(x) = \frac{x^2}{r^2}$$

- Probability density function $f(x) = \frac{2x}{r^2}$
- Expected area of overlap

$$\int_0^r A(x) f(x) dx = \int_0^r \left(2r^2 \cos^{-1}\left(\frac{x}{2r}\right) - x\sqrt{r^2 - \frac{x^2}{4}} \right) \frac{2x}{r^2} dx = 0.5865 \prod r^2$$

  - Expected number of reinforcing neighbours

$$0.5865\, p^2 n^1 = 0.5865 \frac{d^2}{n^1}$$

    - p: probability of sharing sufficient keys
    - n^1: no of expected neighbours
    - d: degree; recall $p = d / n^1$

Effectiveness

- Consider that link has k reinforcing neighbours
  - An adversary must eavesdrop on both the direct link and at least one link of the of the k 2-hop paths
- If the probability of compromising a link is b
  - Probability of breaking the link = $b\ (2b - b^2)^k$
- Communication overhead $\geq 2(0.5865p^2n^1)$
- Studying the Trade-off: $p = 0.33$, $n^1 = 60$, $b = 0.1$
  - 7.66 times additional network traffic
  - Eavesdropping probability is now $6.86 * 10^{-4}$ (from 0.1)

Evaluation

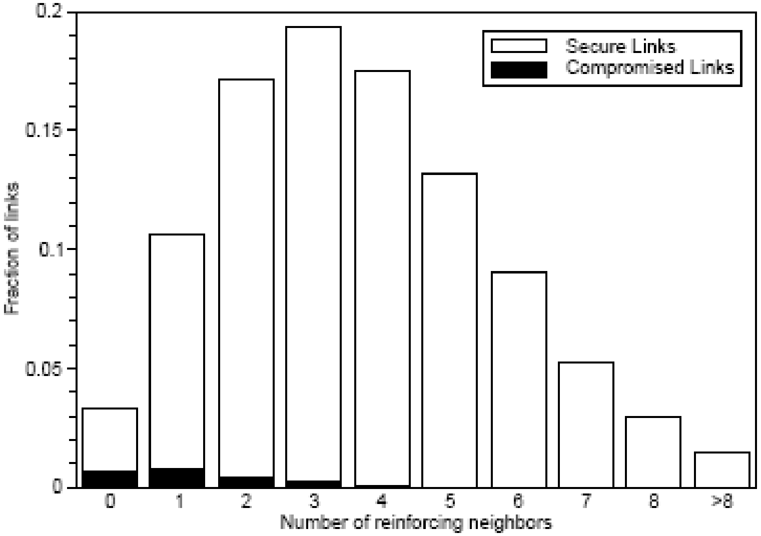
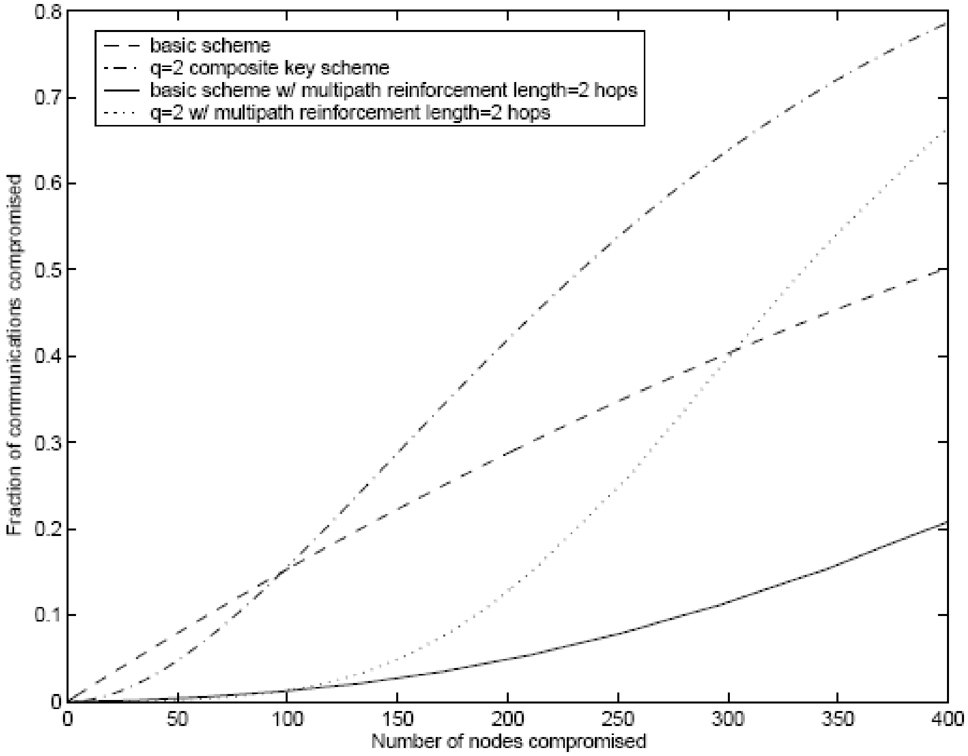
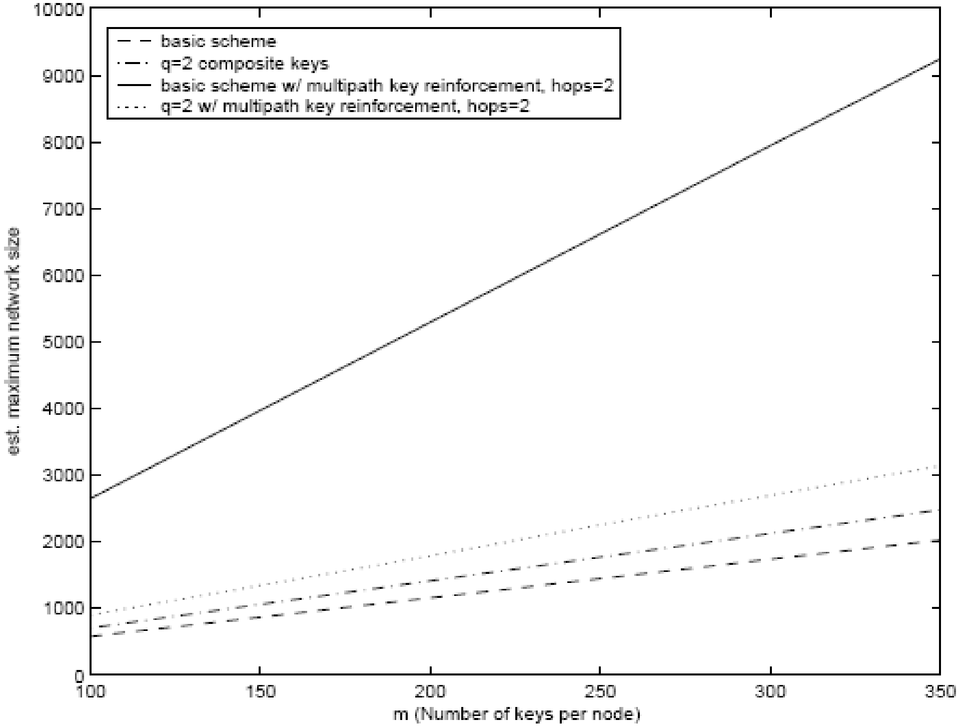


Figure 4. Reinforcement and compromise statistics for base compromise probability $b = 0.2$



Evaluation (2)



- Outperforms q-composite scheme even for q ≥ 2
- Supplementing q - composite scheme with multipath key reinforcement is not a good idea
- Significant boost to network size performance when implemented with the basic scheme
- Can be extended to reinforce path keys

Neither of the two schemes can authenticate the identity of a neighbour

Random Pairwise Keys scheme

- Resilience against node capture
- Node to node identity authentication
- Distributed node revocation
- Resistance to node replication and generation
- Comparable scalability

- A modification of the pairwise keys scheme

Key ring size

- We can calculate the smallest probability *p*, such that the entire network is connected with a high probability *c*
- Therefore each node need store only *np* keys
- Maximum supportable network size $n = m / p$
  - For a key ring size of m
- A node stores the identity of the other node which holds the common key k
- Nodes are certain of the identity since no other nodes can hold k

Initialisation and key setup

- Pre deployment
  - A total of (n = m/p) unique node identities are generated
  - Each node identity is matched with m other randomly selected distinct node IDs
  - Generate a pairwise key for each pair of nodes and store on both key rings along with the node ID of the 'other' node

- Post Deployment
  - Node broadcasts its ID to immediate neighbours
  - Cryptographic handshake is performed to establish neighbours

Multi-hop Range Extension

- Less network traffic and low communication for key discovery
- Increase effective communication range of nodes for key-setup beyond physical communication range
  - Request neighbouring nodes to rebroadcast node ID for k hops
  - Intuitive growth of number of reachable neighbours *x, 4x, 9x, …*
- We have $n = \frac{mn^1}{d}$ Recall $\left[\; p = \frac{d}{n^1};\quad n = \frac{m}{p} \;\right]$
- Susceptibility to DoS attack

Distributed Node Revocation

- Reduce reliance on base station
- Fast response

- Neighbouring nodes broadcast a 'public' vote against the misbehaving node
- Node B severs a communication link with node A on receiving more than a threshold number of public votes against A
- Base station listen in on these broadcasts as well

Properties of the voting scheme

- Compromised nodes cannot revoke arbitrary nodes.
- No voting member of A is able to forge another member’s vote against A.
- Each voting member of A must be able to verify the validity of a broadcast public vote against A.
- Broadcast public votes from one voting member reveal no information that would allow listeners to generate additional public votes.
- Broadcast public votes have no replay value.
- The method of propagating the broadcast to cover the entire network should not be vulnerable to denial of service attack by a malicious node operating within the network.

Revocation Scheme

- Voting members of a node share pairwise keys with the node
- Each voting member of a node A has the following information
  - A random voting key
  - Knowledge of voting key hashes of other (m-1) voting members
- Memory requirement is O(m^2)
- Use of a Merkle tree
  - Store only the root hash
  - Storage space needed to store received votes = $t \log(m)$

Choice of threshold value

- Lower than node degree; large enough to prevent revocation attempts from rogue nodes
- For any of the m keys in a node's key ring, the probability that it is used (the probability another node which has this key is within communication radius) = n^1/n
- Distribution of the degree of a node is the binomial $\left(m, \frac{n^1}{n}\right)$
  - simplifies to $\left(m, \frac{d}{m}\right)$
- We have the mean = *d*, and variance is $d\left(1 - \frac{d}{m}\right)$
  - When d/m is relatively small, heavily skewed to the left
- Expected degree rises slowly with network size
- Threshold must be relatively small (Ex: t ≤ 5, for 1000 ≤ n ≤ 10000)

Discussion of the Revocation Scheme

- No node can have less than *t* neighbours
  - A node with less than *kt* connections must be revoked
    - Degree counting mechanism (later)
- Compromise nodes that shield each other from revocation
  - Compromise nodes around a misbehaving node
  - Present detectable behaviour to utmost (t-1) nodes

- While distributed revocation generates fast response, Base station issued revocations play a necessary role in limiting sophisticated attacks.

Broadcast Mechanism

- Simply rebroadcasting received public nodes leaves the system open to DoS attacks
- Re-broadcast a received and verified public vote for a fixed number of times, at varying intervals
- Every voting member will receive the revocation vote with the same high probability of connectivity of the graph
  - Assume that αn nodes have been deployed, $0.5 \leq \alpha \leq 1$
  - αm voting members have been deployed; each voting member has an expected total of n^1 neighbours within range;
  - Each voting member can find $\left(\frac{\alpha m - 1}{\alpha n - 1}\right) n^1$ voting members in communication range
  - Simplifies to $\frac{mn^1}{n} = d$

Resisting revocation attacks

- Each node can potentially vote against m nodes
    - A significant fraction of total nodes $\left(n = \frac{m}{p}\right)$
  - Compromising a fixed number of nodes could revoke a significant portion of the network

- Restrict by making direct connections a prerequisite
  - Revocation key is stored in a deactivated form k_{Bi}
  - Activation secrets S_{Bi} with the target node B, $0 < i \leq m$
  - To complete key set up, nodes exchange activation secrets
    - Storage requirement: O(m)

Revocation attacks (2)

- Adversary now needs to complete t connections with the target node
- Further, Impose an upper limit d_{max} on the degree of a node
  - Disallow further requests for activation values
  - No of malicious revocation votes restricted to d_{max}
- Strong disincentive to mount a DoS attack via replication for disruption of the network
  - Radio jamming may be a better choice
- Threat not completely eliminated rather less economically viable

Node replication and node generation

- Limit the degree of nodes to d_{max}, a small multiple of d
  - Degree of a node is binomially distributed (m, d/m), heavily skewed to the left
  - d increases slowly with n; almost the order of O(*log* n)
  - d_{max} is generally small compared to m
- Refuse to form connections if d_{max} voting keys are shared
- Degree counting mechanism
  - For every connection between two nodes; broadcast voting keys
  - Each node can track degree of all m nodes that share pairwise keys
  - ***Assumption: A mechanism to store voting keys and verify valid voting keys***
- Storing received votes
  - For small d_{max}, can directly store d_{max}(*log* m) votes
  - Use of merkle trees and m bits to track total number of votes heard
    - Can be further compressed

Evaluation

- Perfect resistance to node capture
- Max supported network size – fixed
- Resistance to revocation attack of distributed scheme
  - Theoretically, attacker can revoke (d_{max} / t) nodes; for every captured node
  - i.e a very small fraction of d
- Revocation attack amplifies the power of an attacker to a small extent
  - Physical destruction
  - Radio Jamming

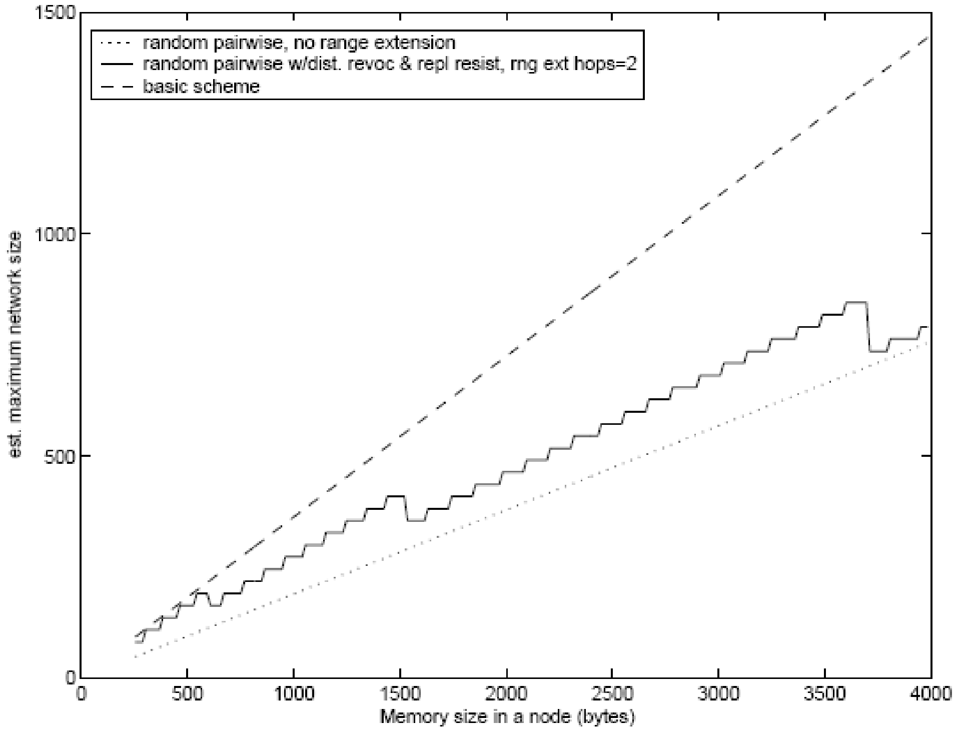


Figure 6. Network sizes for random pairwise key setup compared against the basic scheme with and without multipath key reinforcement. Link keys are 128bits, hash values are 80bits in this simulation. $p = 0.33$, $f_{threshold} = 0.1$

Reviewed related work

- Key distribution for resource starved devices
- Bootstrapping: physical contact with the master device
- Key exchange:
  - asymmetry in computing power
  - An initial secure window for key exchange
- Asymmetric cryptography in ad-hoc networks
- Broadcast encryption

Summarising the three schemes

- Q-composite scheme
  - Improved security under small scale attack vs. greater vulnerability to large scale attack
- 2-hop multipath reinforcement scheme
  - Improved security at cost of communication overhead
  - Deployment density sparse relative to communication radius
- Random Pairwise scheme
  - Security at cost of network size

… .. .

- Any node (new/old/replicated) can try to establish new links within the network
  - Stationary vs. Mobile nodes
  - Longevity of neighbours
  - Window of vulnerability
- Possible motivation for choosing one scheme over the other

Other interesting papers

- Towards a flexible trust establishment framework for sensor networks (Telecommunication Systems 2007)
- “Distributed Detection of Node Replication Attacks in Sensor Networks” (IEEE Security and Privacy Symposium 2005)
- A key management scheme for distributed sensor networks (ACM Conference on Computer and Communication Security, 2002)

Thank You