Evaluating the Vulnerability of Network Traffic Using Joint Security and Routing Analysis

Patrick Tague, *Student Member*, *IEEE*, David Slater, *Student Member*, *IEEE*, Jason Rogers, and Radha Poovendran, *Senior Member*, *IEEE*

Abstract—Joint analysis of security and routing protocols in wireless networks reveals vulnerabilities of secure network traffic that remain undetected when security and routing protocols are analyzed independently. We formulate a class of continuous metrics to evaluate the vulnerability of network traffic as a function of security and routing protocols used in wireless networks. We develop two complementary vulnerability definitions using set theoretic and circuit theoretic interpretations of the security of network traffic, allowing a network analyst or an adversary to determine weaknesses in the secure network. We formalize node capture attacks using the vulnerability metric as a nonlinear integer programming minimization problem and propose the GNAVE algorithm, a **G**reedy **N**ode capture **A**pproximation using **V**ulnerability **E**valuation. We discuss the availability of security parameters to the adversary and show that unknown parameters can be estimated using probabilistic analysis. We demonstrate vulnerability evaluation using the proposed metrics and node capture attacks using the GNAVE algorithm through detailed examples and simulation.

Index Terms—Wireless networks, security, routing, node capture attacks, adversary models.

1 Introduction

Assurance of secure applications and services in wireless networks relies on the properties of *confidentiality* and *integrity*, respectively defined as the ability to keep data secret from unauthorized entities and the ability to verify that data has not been maliciously or accidentally altered [2]. Eschenauer and Gligor recently demonstrated in [3] that these properties can be efficiently compromised by physically capturing network nodes and extracting cryptographic keys from their memories. These *node capture attacks* are possible in most wireless networks due to the unattended operation of wireless nodes and the prohibitive cost of tamper-resistant hardware in portable devices [3]. Furthermore, as shown in [4], an intelligent adversary can improve the efficiency of a node capture attack over that of approaches in recent literature [3], [5], [6], [7] focusing on random node capture using publicly available information leaked from the key assignment protocol.

The aforementioned studies on node capture attacks have all focused on the ability of an adversary to compromise the security of single-hop wireless links. However, messages in a wireless network traverse multiple links and paths between a source and destination node, and a message may be compromised by traversing a single insecure link. The overall security of routed messages is thus dependent on the routing protocol implemented in the wireless network, as well as the physical network topology and the relative positions of the source and destination nodes in the network. Moreover, the fact that a message is transmitted over numerous links between a source and destination node implies that the overall confidentiality and integrity of the routed message may only be as secure as the least secure link, implying that vulnerabilities arise due to the topology of secure links in the wireless network. Hence, the impact of a node capture attack is a function of both the cryptographic protocol which provides link security and the routing protocol which determines the set of links traversed by a given message.

In this article, we introduce a class of metrics to measure the effective security offered in a wireless network as a function of the routing topology and the link security provided by the key assignment protocol. This joint protocol analysis allows a network analyst or an adversary to evaluate the vulnerability of network traffic and isolate weakly secured connections. We approach the problem from an adversarial perspective and show how an intelligent adversary can mount a node capture attack using vulnerability evaluation to focus the attack on the nodes which contribute maximally to the compromise of network traffic. The necessary resource expenditure associated with the node capture attack implies that the optimal attack with minimum resource expenditure corresponds to a minimum cost set of nodes, in contrast to wiretapping attacks in routing or secure network coding [8], [9] which seek a minimum cost set of links. As we show in this article, jointly considering the information from routing and key assignment protocols leads to a significant reduction in resource expenditure in comparison to consideration of information from either protocol separately.

- *P. Tague, D. Slater, and R. Poovendran are with the Network Security Lab (NSL), Department of Electrical Engineering, University of Washington, Paul Allen Center Room AE100R, Campus Box 352500, Seattle, WA 98195-2500. E-mail: {tague, dmslater, rp3}@u.washington.edu.*
- *J. Rogers is with the Center for High Assurance Computer Systems, Information Technology Division, US Naval Research Laboratory, Code 5540, 4555 Overlook Ave. SW, Washington, DC 20375. E-mail: rogers@itd.nrl.navy.mil.*

Manuscript received 4 Jan. 2008; accepted 25 Sept. 2008; published online 28 Oct. 2008.
For information on obtaining reprints of this article, please send e-mail to: tdsc@computer.org, and reference IEEECS Log Number TDSC-2008-01-0003. Digital Object Identifier no. 10.1109/TDSC.2008.60.

1.1 Our Contributions

We make the following contributions in this article:

- We define a class of metrics for the vulnerability of network traffic and formulate the minimum cost node capture attack problem as a nonlinear integer program using the defined vulnerability metrics. We present the GNAVE algorithm, a **G**reedy **N**ode capture **A**pproximation using **V**ulnerability **E**valuation, to approximate the minimum cost node capture attack.
- We provide two complementary realizations for the vulnerability metric by interpreting the compromise of messages using set theoretic and circuit theoretic analogies to evaluate the message security.
- We show that when information about the key assignment protocol is hidden from the adversary using privacy-preserving protocols, the indeterminate quantities can be estimated probabilistically without significant degradation in the attack performance.
- We demonstrate the impact of node capture attacks using the GNAVE algorithm in wireless networks with examples of both classical routing and network coding protocols. Furthermore, we compare the resource expenditure required for node capture attacks using the GNAVE algorithm to previously proposed strategies via simulation.

The remainder of this article is organized as follows: In Section 2, we state the assumed wireless network, key assignment, and adversary models. In Section 3, we define the class of vulnerability metrics for routed traffic, formulate the minimum cost node capture attack problem, and propose the GNAVE algorithm using a greedy heuristic for node capture. In Section 4, we provide two realizations of the vulnerability metrics using set and circuit theoretic formulations. In Section 5, we show how to estimate parameters that are unknown due to the use of privacy-preserving protocols. In Section 6, we provide examples and simulation of node capture attacks using vulnerability evaluation. In Section 7, we state our conclusions.

2 Models and Notation

In this section, we state the assumed wireless network, key assignment, and adversary models. We summarize the notation used throughout this article in Table 1.

2.1 Network Model

The topology of the wireless network with a set of nodes $\mathcal{N}$ is represented by the directed *network graph* $G = (\mathcal{N}, L)$. The link set L contains all ordered pairs of one-hop communicating neighbors, equivalent to an asymmetric relation [10], such that (i, j) is in L for $i \neq j$ if and only if node i can reliably send messages to node j without intermediate relay nodes. The link set L is dependent on parameters such as node location and configuration and properties of the radios, transmission medium, and MAC layer protocols.

We denote the subsets of $\mathcal{N}$ of message source and destination nodes in the network as $\mathcal{S}$ and $\mathcal{D}$, respectively. The set of source-destination pairs is denoted $\mathcal{T} \subseteq \mathcal{S} \times \mathcal{D}$ and is constructed based on the routing protocol decisions.

TABLE 1
A Summary of Notation is Provided for Reference

Symbol	Definition
$\mathcal{N}$	Set of N wireless nodes
L	Set of ordered pairs of one-hop neighbor nodes
G	Network graph $(\mathcal{N}, L)$
$\mathcal{K}, \mathcal{L}$	Set of keys, labels
$\mathcal{K}_i, \mathcal{L}_i$	Set of keys, labels assigned to node $i \in \mathcal{N}$
$\mathcal{K}_{ij}, \mathcal{L}_{ij}$	Set of keys, labels shared by nodes i and j
$\mathcal{S}, \mathcal{D}$	Set of source, destination nodes
$\mathcal{T}$	Subset of $\mathcal{S} \times \mathcal{D}$ of source-destination pairs
$\mathcal{R}_{sd}$	Set of paths forming the route from s to d
G_{sd}	Route subgraph of G corresponding to $\mathcal{R}_{sd}$
f_π	Fraction of $\mathcal{R}_{sd}$ traffic traversing π
$\mathcal{K}_{sd}^E$	Set of keys securing the end-to-end link (s, d)
$\mathcal{T}_A$	Adversary's target subset of $\mathcal{T}$
$\mathcal{C}$	Subset of $\mathcal{N}$ of captured nodes
$\mathcal{K}_\mathcal{C}, L_\mathcal{C}$	Set of compromised keys, links when $\mathcal{C}$ captured
w_i	Weight or cost of capturing node $i \in \mathcal{N}$
ρ_{sd}	Weight representing adversary's route preference
$V_{sd}(\mathcal{C})$	Route vulnerability of $\mathcal{R}_{sd}$ when $\mathcal{C}$ captured
$\nu_i(\mathcal{C})$	Incremental value of node i when $\mathcal{C}$ captured
$R_\mathcal{C}(i, j)$	Link resistance of (i, j) when $\mathcal{C}$ captured
$R_\mathcal{C}(\mathcal{R}_{sd})$	Route resistance of $\mathcal{R}_{sd}$ when $\mathcal{C}$ captured

For a given source-destination pair $(s, d) \in \mathcal{T}$, the routing protocol will construct one or more directed routing *paths* through G, where a path is defined as a set of sequential links in L. We define the *route* $\mathcal{R}_{sd}$ as the set of all paths traversed by any message from s to d, and we let f_π denote the fraction of traffic from s to d that traverses the given path $\pi \in \mathcal{R}_{sd}$. The route $\mathcal{R}_{sd}$ can be represented graphically by the *route subgraph* G_{sd} of G consisting of nodes and directed links traversed by at least one routing path $\pi \in \mathcal{R}_{sd}$ from s to d.

We define the following classes of routing protocols, partitioning the space of routing protocols based on the dependence of messages routed along different (not necessarily disjoint) paths, as follows:

Definition 1. *The class of* independent path routing *protocols consists of any protocol which uses one or more paths to route separate messages such that messages traversing different paths are independently coded and secured.*

The class of independent path routing protocols contains, for example, protocols using a single, fixed path such as AODV [11] or DSR [12] as well as protocols using multiple paths such as GBR [13] or GEAR [14]. The route $\mathcal{R}_{sd}$ under independent path routing is equivalent to the superposition of $|\mathcal{R}_{sd}|$ single-path routes, where each single-path route $\{\pi\}$ for $\pi \in \mathcal{R}_{sd}$ is weighted by the corresponding traffic fraction f_π.

Definition 2. *The class of* dependent path routing *protocols consists of any protocol which uses multiple paths in which packets traversing separate paths are jointly coded, fragmented, or secured.*

The class of dependent path routing protocols contains, for example, protocols based on threshold secret sharing [15] and network coding [8], [9], [16] in which a set of coded

packets must be jointly decoded in order to recover the original set of messages.

2.2 Key Assignment Model

We assume the existence of a secure key assignment mechanism as follows: Let $\mathcal{K}$ be a set of symmetric cryptographic keys and $\mathcal{L}$ be a corresponding set of publicly available key labels. Each node $i \in \mathcal{N}$ is assigned a subset $\mathcal{K}_i \subseteq \mathcal{K}$ and the corresponding subset $\mathcal{L}_i \subseteq \mathcal{L}$. We denote the subset of keys shared by nodes i and j as $\mathcal{K}_{ij} = \mathcal{K}_i \cap \mathcal{K}_j$ and allow communication between i and j if and only if $\mathcal{K}_{ij} \neq \emptyset$.[1] We assume that nodes i and j use the entire set $\mathcal{K}_{ij}$ of shared keys to secure the link (i, j), so the strength of the link security is directly related to the number of shared keys. We assume that nodes i and j compute the intersection $\mathcal{L}_{ij} = \mathcal{L}_i \cap \mathcal{L}_j$ in order to determine the set of shared keys $\mathcal{K}_{ij}$ using a protocol from one of the following classes.

Definition 3. *The class of* public label exchange *protocols consists of any protocol which provides necessary information for any node $j \in \mathcal{N}$ to compute the set $\mathcal{L}_i$ of key labels for any node $i \in \mathcal{N}$.*

The class of public label exchange protocols contains such protocols as the public broadcast of $\mathcal{L}_i$ by each node $i \in \mathcal{N}$ as in [3] or the use of a public identity-based function to compute $\mathcal{L}_i$ as a function of i as in [17].

Definition 4. *The class of* privacy-preserving set intersection *protocols consists of any protocol which provides necessary information for any node $j \in \mathcal{N}$ to only compute the set $\mathcal{L}_{ij}$ of key labels shared with any node $i \in \mathcal{N}$ without giving any information to j about the remaining key labels in $\mathcal{L}_i \setminus \mathcal{L}_j$.*

The class of privacy-preserving set intersection protocols contains such protocols as the challenge-response protocol proposed in [3] in which each node $i \in \mathcal{N}$ computes a random nonce α and broadcasts α and the challenge $E_k(\alpha)$ for each $k \in \mathcal{K}_i$.

In addition to the link security provided by the set of shared keys $\mathcal{K}_{ij}$ for each link (i, j), we consider the incorporation of an additional *end-to-end* security mechanism for each route $\mathcal{R}_{sd}$ which depends only on the source s and destination d. If it is physically possible and allowed by policy, the source node s can compute the set $\mathcal{K}_{sd}$ of keys shared with the destination node d and additionally secure messages in the route $\mathcal{R}_{sd}$ using the shared keys $\mathcal{K}_{sd}$. We denote the set of keys securing the end-to-end connection between s and d as $\mathcal{K}_{sd}^E$, noting that $\mathcal{K}_{sd}^E = \mathcal{K}_{sd}$ if s and d are able and allowed to use end-to-end security and $\mathcal{K}_{sd}^E = \emptyset$ otherwise. We include the additional end-to-end secure link (s, d) in the route subgraph G_{sd} with the corresponding link security depending only on $\mathcal{K}_{sd}^E$.

2.3 Adversarial Model

We consider a polynomial-time adversary with the ability and resources to eavesdrop on and record messages throughout the network, capture nodes, and extract cryptographic keys from the memory of captured nodes. We assume that the adversary has knowledge of the key assignment and routing protocols, including protocol parameters, and can participate actively in any network protocols by assuming the roles of captured, replicated, or fabricated nodes. We further assume that the route subgraph G_{sd} for each $(s, d) \in \mathcal{T}$ is available to the adversary or is computable using traffic analysis and estimation [18].

The primary goal of the adversary is to compromise the confidentiality and integrity of all messages routed between a *target set* of source-destination pairs denoted $\mathcal{T}_A \subseteq \mathcal{T}$ by extracting cryptographic keys from the memory of captured nodes $\mathcal{C} \subseteq \mathcal{N}$ with minimum resource expenditure. The adversary thus captures nodes intelligently by associating an individual weight or *cost* w_i with the resource expenditure required to capture each node $i \in \mathcal{N}$, as in [4]. We do not address further attacks on network protocols and services that can be performed as a result of message compromise.

3 Route Vulnerability Metrics under Node Capture Attacks

In this section, we define a class of route vulnerability metrics (RVMs) to quantify the effective security of traffic traversing a given route $\mathcal{R}_{sd}$. Using the RVM definition, we formulate the minimum cost node capture attack problem as a nonlinear integer programming minimization problem. Since determining the optimal node capture attack is likely infeasible, we propose the GNAVE algorithm using a greedy heuristic to iteratively capture nodes which maximize the increase in route vulnerability.

3.1 Route Vulnerability Metric (RVM)

In order to evaluate the effect of a node capture attack on the effective security of traffic traversing a route $\mathcal{R}_{sd}$, we formally define link, path, and route compromise due to the capture of a subset $\mathcal{C} \subseteq \mathcal{N}$ of network nodes. We denote the set of keys recovered by the adversary in capturing the subset $\mathcal{C}$ as $\mathcal{K}_\mathcal{C} = \bigcup_{i \in \mathcal{N}} \mathcal{K}_i$. If a message traverses a link which is secured by keys in $\mathcal{K}_\mathcal{C}$, the security of the message is compromised. The compromise of individual links in the network, with respect to the network and routing models in Section 2, is defined as follows:

Definition 5. *The link $(i, j) \in L$ or $(s, d) \in \mathcal{T}$ is compromised if and only if $\mathcal{K}_{ij} \subseteq \mathcal{K}_\mathcal{C}$ or $\mathcal{K}_{sd}^E \subseteq \mathcal{K}_\mathcal{C}$, respectively, and the set of all compromised links is denoted $L_\mathcal{C} \subseteq L \cup \mathcal{T}$.*

Using Definition 5, we further define the compromise of paths and message routes as follows:

Definition 6. *The path $\pi \in \mathcal{R}_{sd}$ is compromised if and only if $(s, d) \in L_\mathcal{C}$ and there is at least one link (i, j) in π for which $(i, j) \in L_\mathcal{C}$.*

Note that the inclusion of the end-to-end link (s, d) in the requirement for path compromise indicates that any message traversing a compromised path can be eavesdropped or modified by the adversary.

1. This requirement can be strengthened as in [5] to require $|\mathcal{K}_{ij}| \geq q$ for a fixed integer $q \geq 1$, though we do not explicitly address this requirement.

Definition 7. *The route $\mathcal{R}_{sd}$ for $(s,d) \in \mathcal{T}$ is compromised if and only if every path $\pi \in \mathcal{R}_{sd}$ is compromised.*

Using Definition 7, an adversary can compute the fraction of target routes compromised due to the capture of a set of nodes $\mathcal{C}$. However, this evaluation does not provide the adversary with a method for selection of the set $\mathcal{C}$. Furthermore, the fraction of compromised target routes does not provide any indication of the contribution of nodes in $\mathcal{C}$ toward the future compromise of additional routes, as the compromise of a route is a binary event.

To adequately capture the progression toward the compromise of additional routes, we introduce the metric of route vulnerability $V_{sd}(\mathcal{C})$ as defined by the following RVM class.

Definition 8. *The* route vulnerability *$V_{sd}(\mathcal{C})$ of the route $\mathcal{R}_{sd}$ due to the capture of nodes in $\mathcal{C}$ is defined as any of the class of functions mapping into the unit interval $[0,1]$ such that*

1. *$V_{sd}(\emptyset) = 0$, where $\emptyset$ is the empty set,*
2. *$V_{sd}(\mathcal{C}) = 1$ if and only if $\mathcal{R}_{sd}$ is compromised when $\mathcal{C}$ is captured, and*
3. *$0 < V_{sd}(\mathcal{C}) < 1$ if and only if $\mathcal{R}_{sd}$ is not compromised when $\mathcal{C}$ is captured but $\mathcal{C}$ contributes to the weakening of the security of at least one link in the route $\mathcal{R}_{sd}$.*

The class of RVMs thus relaxes the binary notion of route compromise to a continuous measure of the progress of the attack and allows for comparison of partial compromise by different sets $\mathcal{C}_1$ and $\mathcal{C}_2$ of captured nodes.

3.2 Node Capture Attack Formulation

For any RVM realization satisfying the conditions of Definition 8, we devise a node capture strategy that maximizes the progression toward the goal of compromising all routes $\mathcal{R}_{sd}$ for $(s,d) \in \mathcal{T}_A$. The choice of subset $\mathcal{C}$ requiring the minimum resource expenditure is thus given by the following minimum cost node capture problem.

Problem: Minimum Cost Node Capture Attack
Given: $\mathcal{L}_i$, w_i for $i \in \mathcal{N}$, $\mathcal{R}_{sd}$ for $(s,d) \in \mathcal{T}_A$
Find: $\mathcal{C} \subseteq \mathcal{N}$

such that $\sum_{i \in \mathcal{C}} w_i$ is minimized
and $V_{sd}(\mathcal{C}) = 1$ for all $(s,d) \in \mathcal{T}_A$.

In general, based on Definition 6 of path compromise, the metric $V_{sd}(\mathcal{C})$ is nonlinear in the entries of $\mathcal{C}$. Hence, the minimum cost node capture attack above is a nonlinear integer programming minimization problem, known to be NP-hard [10], [19]. We thus propose the use of a greedy heuristic that iteratively adds nodes to $\mathcal{C}$ based on maximizing the increase in route vulnerability $V_{sd}(\mathcal{C})$ at each step. The heuristic is thus similar to a known greedy heuristic for set covering [20] and linear integer programming [19]. However, due to the nonlinearity in $V_{sd}(\mathcal{C})$, the worst-case performance of the greedy heuristic cannot be analyzed using the ratio-bound analysis in [10], [19], [20] and is left as an open problem.

To maximize the route vulnerability $V_{sd}(\mathcal{C})$ with minimum resource expenditure, it is beneficial to the adversary to attempt to maximize the vulnerability resulting from the capture of each individual node using the information recovered from previously captured nodes. The contribution of a node i is thus given by the increase in route vulnerability $V_{sd}(\mathcal{C} \cup \{i\}) - V_{sd}(\mathcal{C})$ due to the addition of i to $\mathcal{C}$. Allowing for an additional weight ρ_{sd} to indicate the adversary's preference to compromise the route $\mathcal{R}_{sd}$ over other routes, the value of each node i is defined as follows:

Definition 9. *The individual* incremental node value *of adding node $i \in \mathcal{N}$ to $\mathcal{C}$ is defined as*

$$\nu_i(\mathcal{C}) = \sum_{(s,d) \in \mathcal{T}_A} \rho_{sd}(V_{sd}(\mathcal{C} \cup \{i\}) - V_{sd}(\mathcal{C}))$$

for any route vulnerability function $V_{sd}(\mathcal{C})$ satisfying the conditions in Definition 8.

To maximize the cost effectiveness of the node capture attack at each iteration, the adversary chooses to capture the node with maximum incremental value per unit cost $\nu_i(\mathcal{C})/w_i$. Based on this greedy approach, we propose the GNAVE algorithm as follows:

GNAVE Algorithm
Given: $\mathcal{L}_i$, w_i for $i \in \mathcal{N}$, $\mathcal{R}_{sd}$ for $(s,d) \in \mathcal{T}_A$
$\mathcal{C} \leftarrow \emptyset$
while there exists $(s,d) \in \mathcal{T}_A$ with $V_{sd}(\mathcal{C}) < 1$ **do**
  $i^* \leftarrow \underset{i \in \mathcal{N}}{\operatorname{argmax}}\, \nu_i(\mathcal{C})/w_i$
  $\mathcal{C} \leftarrow \mathcal{C} \cup \{i^*\}$
end while

We note that the GNAVE algorithm being greedy implies that the attack performance depends only on the order of the weighted node values $\nu_i(\mathcal{C})/w_i$ for the nodes $\mathcal{N} \setminus \mathcal{C}$. In order to illustrate the effect of node capture attacks using the GNAVE algorithm, we next provide candidate realizations of the RVM $V_{sd}(\mathcal{C})$.

4 RVM Realizations

In this section, we propose two RVM realizations satisfying the conditions in Definition 8, noting that there is a high degree of freedom in the given conditions. We present each RVM realization for each of the routing protocol classes discussed in Section 2.1, hereafter denoting the route vulnerability for independent and dependent path routing protocols as $V_{sd}^I(\mathcal{C})$ and $V_{sd}^D(\mathcal{C})$, respectively. The definitions presented in this section are derived using the following necessary and sufficient condition for the compromise of a route $\mathcal{R}_{sd}$ with respect to the edge cuts [10] of the route subgraph G_{sd}.

Theorem 1. *The route $\mathcal{R}_{sd}$ is compromised if and only if the set $L_{\mathcal{C}}$ of compromised links contains at least one (s,d) edge cut of the route subgraph G_{sd} as a subset.*

Proof. To prove the forward implication, suppose that $\mathcal{R}_{sd}$ is compromised. By Definitions 6 and 7, there is at least one compromised link (i_π, j_π) in each path $\pi \in \mathcal{R}_{sd}$ and the end-to-end link (s,d) is compromised. Let $L_{\text{cut}} = \{(i_\pi, j_\pi) : \pi \in \mathcal{R}_{sd}\} \subseteq L_{\mathcal{C}}$. Since each path π traverses at least one edge in L_{cut}, $L_{\text{cut}} \cup \{(s,d)\}$ is an edge cut of G_{sd}.

To prove the reverse implication, let L_{cut} be an edge cut of G_{sd}. By the definition of an edge cut, $(s,d) \in L_{\text{cut}}$

and each path π from s to d in G_{sd} traverses at least one link in L_{cut}. Hence, by Definition 6, every path π in $\mathcal{R}_{sd}$ is compromised, implying by Definition 7 that the route itself is compromised. □

Theorem 1 thus implies that the task of compromising each route $(s,d) \in \mathcal{T}_A$ is equivalent to capturing a set of nodes $\mathcal{C}$ leading to the compromise of an edge cut of G_{sd}. We thus formulate two RVM realizations using the properties of edge cuts of G_{sd}.

4.1 Set Theoretic Version of RVM

We formulate a set theoretic RVM realization $V_{sd}(\mathcal{C})_{\text{SET}}$ by interpreting the properties of edge cuts of G_{sd} set theoretically. From Theorem 1, the existence of a compromised edge cut set $L_{\text{cut}} \subseteq L_{\mathcal{C}}$ of the route subgraph G_{sd} implies that the route $\mathcal{R}_{sd}$ is compromised. In terms of the set $\mathcal{K}_{\mathcal{C}}$ of compromised keys, a necessary and sufficient condition for $L_{\mathcal{C}}$ to contain an edge cut set of G_{sd} is

$$\mathcal{K}_{sd} \subseteq \mathcal{K}_{\mathcal{C}} \text{ and } \forall \pi \in \mathcal{R}_{sd}, \exists (i,j) \in \pi, \mathcal{K}_{ij} \subseteq \mathcal{K}_{\mathcal{C}}.$$

Letting $\mathbf{1}(\cdot)$ denote the binary indicator function of a specified event, Theorem 1 thus implies that the first two conditions of Definition 8 can be satisfied by defining a binary RVM equal to

$$\mathbf{1}(\mathcal{K}_{sd} \subseteq \mathcal{K}_{\mathcal{C}}) \prod_{\pi \in \mathcal{R}_{sd}} \left(1 - \prod_{(i,j)\in\pi} \left(1 - \mathbf{1}(\mathcal{K}_{ij} \subseteq \mathcal{K}_{\mathcal{C}})\right) \right). \quad (1)$$

However, this function does not satisfy the third condition of Definition 8 as the resulting function does not take continuous values between 0 and 1.

The above formulation provides insight into the route vulnerability, however, suggesting that a valid RVM can be obtained with minor modifications. First, to ensure that any compromised path is accounted for in the vulnerability evaluation, the product over all paths in $\mathcal{R}_{sd}$ can be replaced by a weighted summation over the corresponding paths, including the secure end-to-end link (s,d) as a single-hop path. We denote the relative weight assigned to the secure end-to-end link (s,d) as f_{sd} with the assumption that $f_{sd} > 0$ is allowed to vary arbitrarily when the additional end-to-end secure link is used and that $f_{sd} = 0$ otherwise, thus impacting the choice of captured nodes. We relax the binary condition imposed by the indicator function $\mathbf{1}(\mathcal{K}_{ij} \subseteq \mathcal{K}_{\mathcal{C}})$ by the function $\phi_{ij}(\mathcal{C})$ equal to the fraction of keys in $\mathcal{K}_{ij}$ that are contained in $\mathcal{K}_{\mathcal{C}}$, given by

$$\phi_{ij}(\mathcal{C}) = \begin{cases} \frac{|\mathcal{K}_{ij} \cap \mathcal{K}_{\mathcal{C}}|}{|\mathcal{K}_{ij}|}, & \text{if } \mathcal{K}_{ij} \neq \emptyset, \\ 1, & \text{otherwise} \end{cases} \quad (2)$$

for links in L and

$$\phi_{sd}(\mathcal{C}) = \begin{cases} \frac{|\mathcal{K}^E_{sd} \cap \mathcal{K}_{\mathcal{C}}|}{|\mathcal{K}^E_{sd}|}, & \text{if } \mathcal{K}^E_{sd} \neq \emptyset, \\ 1, & \text{otherwise} \end{cases} \quad (3)$$

for the secure end-to-end link (s,d). Applying this relaxation to the right-hand side of (1) thus yields the following RVMs for independent and dependent path routing protocols, which vary only in the weighting of individual paths in $\mathcal{R}_{sd}$.

For independent path routing protocols, the compromise of an individual path $\pi \in \mathcal{R}_{sd}$ is sufficient to allow the adversary to recover a fraction f_π of the traffic from s to d. Applying the continuous relaxation to the right-hand side of (1) for each single path route in $\mathcal{R}_{sd}$ and summing over the single path routes with corresponding weights f_π, including the end-to-end link (s,d) with weight f_{sd}, yields the RVM for independent path routing protocols as

$$V^I_{sd}(\mathcal{C})_{\text{SET}} = \frac{f_{sd}\phi_{sd}(\mathcal{C}) + 1}{1 + f_{sd}} - \sum_{\pi \in \mathcal{R}_{sd}} \frac{f_\pi}{1 + f_{sd}} \prod_{(i,j)\in\pi} \left(1 - \phi_{ij}(\mathcal{C})\right). \quad (4)$$

For dependent path routing protocols, even though the compromise of an individual path does not reveal any information to the adversary, it brings the adversary closer to compromising the route. Hence, we obtain the corresponding RVM by applying the continuous relaxation to the right-hand side of (1) and summing over the equally weighted single path routes, including the end-to-end link (s,d) with weight f_{sd}, yielding

$$V^D_{sd}(\mathcal{C})_{\text{SET}} = \frac{f_{sd}\phi_{sd}(\mathcal{C}) + 1}{1 + f_{sd}} - \frac{1}{|\mathcal{R}_{sd}|(1 + f_{sd})} \sum_{\pi \in \mathcal{R}_{sd}} \prod_{(i,j)\in\pi} \left(1 - \phi_{ij}(\mathcal{C})\right). \quad (5)$$

The set theoretic formulation of the RVM $V_{sd}(\mathcal{C})_{\text{SET}}$ in this section is derived by explicitly analyzing the necessary condition for the existence of an edge cut of G_{sd}. In what follows, we provide an alternate approach which jointly considers all edge cuts of G_{sd}.

4.2 Circuit Theoretic Version of RVM

We formulate a circuit theoretic RVM realization $V_{sd}(\mathcal{C})_{\text{CIR}}$ which jointly considers all edge cuts of G_{sd} using duality properties of planar graphs and electric circuits. As a basis of the formulation, we first outline a mapping between edge cuts of G_{sd} and current flowing through a corresponding electric circuit $\mathcal{E}_{sd}$.

4.2.1 Mapping Edge Cuts to Current Flow When G_{sd} Is a Planar Graph

We provide a mapping between the joint evaluation of all edge cuts of the route subgraph G_{sd} and the resistance of an electric circuit when G_{sd} is a planar graph [21], i.e., it is possible to draw G_{sd} with no edges intersecting. The mapping is formulated by mapping a single edge cut L_{cut} of G_{sd} to a single current path through an electric circuit $\mathcal{E}_{sd}$. Fig. 1 provides an illustration of each step of the mapping.

Step 1. The edge cut L_{cut} is mapped to a continuous, directed curve z_L which *crosses* G_{sd} [21], crossing the edges in L_{cut} in a direction perpendicular to each edge. Since the graph G_{sd} is directed, the edge (i,j) only appears in the edge cut L_{cut} if i is on the source-side of the cut and j is on the destination-side of the cut. As an example, consider the

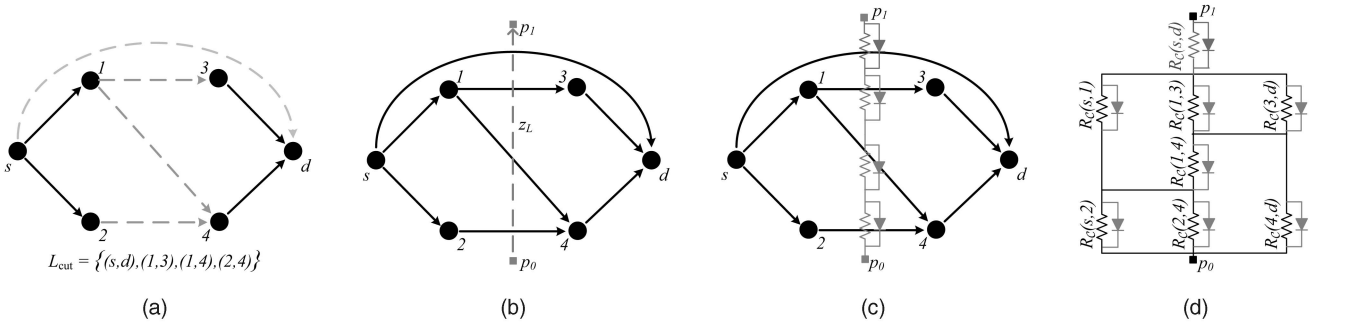

Fig. 1. The mapping from an edge cut L_{cut} of the route subgraph G_{sd} to a current path through the electric circuit $\mathcal{E}_{sd}$ is illustrated. (a) An edge cut L_{cut} of the route subgraph G_{sd} is illustrated. (b) The edge cut is mapped to curve z_L directed from p_0 to p_1 and crossing the edges in L_{cut}. (c) The curve z_L is replaced by a wire, and a resistor-diode pair is inserted at each point where the curve z_L crosses an edge in G_{sd}. (d) The circuit $\mathcal{E}_{sd}$ is illustrated by combining the wires and resistors for each possible edge cut L_{cut}. The diode in parallel with each resistor accounts for the orientation of edges in G_{sd}.

edge cut $L = \{(s,d),(s,1),(4,d)\}$ of G_{sd} in Fig. 1a. If the direction of the edge (1, 4) is ignored, L is no longer an edge cut, as the path $\{(s,2),(2,4),(4,1),(1,3),(3,d)\}$ is not compromised. The edge cut L_{cut} in Fig. 1a is thus mapped to the curve z_L in Fig. 1b.

Step 2. The curve z_L crossing G_{sd} is mapped to a wire carrying electric current from the starting point p_0 to the ending point p_1. To represent the cost associated with the capture of nodes in $\mathcal{C}$ to compromise the edge cut L_{cut}, a resistor of resistance $R_{\mathcal{C}}(i,j)$ is inserted at the point in the wire where the curve z_L crosses the edge $(i,j) \in L_{\text{cut}}$. To maintain edge directionality, an ideal diode is inserted in parallel with the resistor to zero the resistance in the opposite direction. The curve z_L in Fig. 1b is thus mapped to the resistive current path from p_0 to p_1 in Fig. 1c.

Step 3. The resistive current paths corresponding to the edge cuts L_{cut} of the graph G_{sd} are then combined into an electric circuit $\mathcal{E}_{sd}$ with a resistor of resistance $R_{\mathcal{C}}(i,j)$ and the corresponding diode corresponding to each edge (i,j) in G_{sd}. The route subgraph G_{sd} in Fig. 1a is thus mapped to the circuit $\mathcal{E}_{sd}$ in Fig. 1d, consisting of the composition of all current paths from p_0 to p_1 that cross G_{sd}. The resistance to the current along each current path in $\mathcal{E}_{sd}$ corresponds to the difficulty faced in compromising the corresponding edge cut, so the equivalent resistance of $\mathcal{E}_{sd}$ is proportional to the strength of the message security for the route $\mathcal{R}_{sd}$.

By construction, the underlying graph structure of the circuit $\mathcal{E}_{sd}$ is related to the graph G_{sd} using the planar dual [21], with an auxiliary edge (d,s) to close the directed network flow. Hence, the mapping provides a one-to-one correspondence between the directed edges in G_{sd} and the resistor-diode pairs in $\mathcal{E}_{sd}$, implying that the circuit $\mathcal{E}_{sd}$ can be constructed from the route subgraph G_{sd} without explicitly computing the edge cuts of G_{sd}.

4.2.2 Mapping Edge Cuts to Current Flow When G_{sd} Is a Nonplanar Graph

We provide a mapping between the joint evaluation of all edge cuts of the route subgraph G_{sd} and the resistance of an electric circuit when G_{sd} is not a planar graph.

We construct a second electric circuit $\mathcal{E}^*_{sd}$ using duality properties of electric circuits [22] which state that the behavior of circuit elements in one circuit mirror that of the corresponding elements of the dual circuit. The dual circuit $\mathcal{E}^*_{sd}$ is constructed from $\mathcal{E}_{sd}$ by replacing series connections with parallel connections, replacing current loops with voltage nodes, replacing impedance quantities with admittance quantities, and vice versa. As shown in [22], once the dual circuit $\mathcal{E}^*_{sd}$ is constructed, properties of $\mathcal{E}_{sd}$ can be observed by transforming the corresponding properties of the dual circuit $\mathcal{E}^*_{sd}$. In particular, the effective resistance of $\mathcal{E}_{sd}$ is equal to the inverse of the effective resistance of $\mathcal{E}^*_{sd}$.

We note that the aforementioned interchange of current loops and voltage nodes implies that the underlying graphs of the dual circuits $\mathcal{E}_{sd}$ and $\mathcal{E}^*_{sd}$ are planar duals of each other. Since the planar dual of the graph corresponding to $\mathcal{E}_{sd}$ is the route subgraph, G_{sd}, the dual circuit $\mathcal{E}^*_{sd}$ can be constructed directly from G_{sd} by replacing each directed edge in G_{sd} with a resistor and diode in series. Using circuit duality, the resistance of the resistor in $\mathcal{E}^*_{sd}$ corresponding to the edge (i,j) in G_{sd} is labeled with the resistance $R_{\mathcal{C}}(i,j)^{-1}$. Moreover, the inverse of the equivalent resistance of the dual circuit $\mathcal{E}^*_{sd}$ is equal to the equivalent resistance of the circuit $\mathcal{E}_{sd}$. To illustrate the dual circuit construction, we provide Fig. 2 as the dual $\mathcal{E}^*_{sd}$ corresponding to the nonplanar graph obtained by adding the edge (2, 3) to the graph G_{sd} in Fig. 1a.

Since $\mathcal{E}^*_{sd}$ is constructed directly from G_{sd}, this technique does not rely on the planarity of G_{sd}, thus providing an extension to nonplanar route subgraphs. Given the above mapping, we next show how the equivalent resistances of the circuit $\mathcal{E}_{sd}$ and the dual circuit $\mathcal{E}^*_{sd}$ can be used to define the route vulnerability $V_{sd}(\mathcal{C})_{\text{CIR}}$.

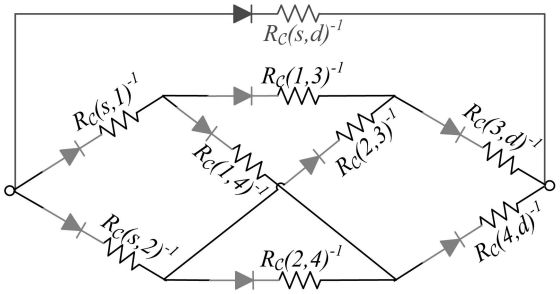

Fig. 2. The route subgraph G_{sd} in Fig. 1a is made nonplanar by adding the edge (2, 3). The dual circuit $\mathcal{E}^*_{sd}$ corresponding to G_{sd} is illustrated. Note that the parallel resistor-diode pair in $\mathcal{E}_{sd}$ is transformed to a series resistor-diode pair in $\mathcal{E}^*_{sd}$.

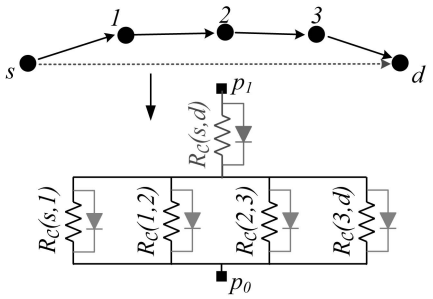

Fig. 3. The resistance $R_C(\{\pi\})$ of the single path route π is given by (6) as the equivalent resistance of the parallel resistors $R_C(i,j)$ added to the link resistance of the end-to-end secure link (s,d).

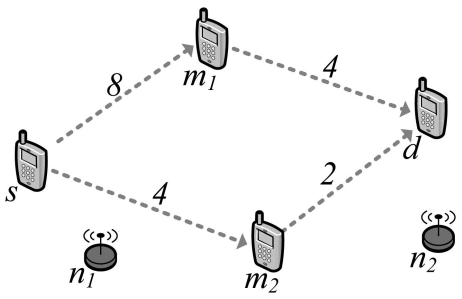

Fig. 4. We consider two examples using the illustrated network with source s, intermediate nodes m_1 and m_2, and destination d. Each link is labeled with the number of keys securing the link.

4.2.3 Evaluating Vulnerability with Circuit Analysis

We provide a definition for the route vulnerability $V_{sd}(\mathcal{C})_{\text{CIR}}$ using the circuit theoretic mappings in Sections 4.2.1 and 4.2.2. We first provide a definition of the resistance values $R_C(i,j)$ used in the circuits $\mathcal{E}_{sd}$ and $\mathcal{E}^*_{sd}$ with respect to the key assignment parameters in Section 2.2.

Definition 10. *The* link resistance $R_C(i,j)$ *of the resistors in* $\mathcal{E}_{sd}$ *and* $\mathcal{E}^*_{sd}$ *is equal to the number of keys securing the link* (i,j) *that are not compromised, given by*

$$R_C(i,j) = |\mathcal{K}_{ij} \setminus \mathcal{K}_C|$$

for links in L *and*

$$R_C(s,d) = |\mathcal{K}^E_{sd} \setminus \mathcal{K}_C|$$

for the secure end-to-end link (s,d).

We note that the link resistance values are a measure of the residual security of individual links to the capture of nodes in $\mathcal{C}$. The following definition extends this concept to the entire route subgraph G_{sd}.

Definition 11. *The* route resistance $R_C(\mathcal{R}_{sd})$, *quantifying the resilience of the route* $\mathcal{R}_{sd}$ *to the capture of nodes in* $\mathcal{C}$, *is defined as the inverse of the equivalent resistance of the dual circuit* $\mathcal{E}^*_{sd}$. *By the duality property,* $R_C(\mathcal{R}_{sd})$ *is equal to the equivalent resistance of the circuit* $\mathcal{E}_{sd}$ *when* G_{sd} *is planar.*

We note that the link resistance is a function only of the key assignment protocol, while the route resistance is a function of both the key assignment and routing protocols.

We next define the route vulnerability $V_{sd}(\mathcal{C})_{\text{CIR}}$ as a function of the route resistance $R_C(\mathcal{R}_{sd})$. Since the overall resistances of the circuits $\mathcal{E}_{sd}$ and $\mathcal{E}^*_{sd}$ are, respectively, inversely and directly proportional to the adversary's ability to compromise the route $\mathcal{R}_{sd}$, the route vulnerability is defined to be proportional to the inverse $(1+R_C(\mathcal{R}_{sd}))^{-1}$, scaled by a function of the initial condition $R_\emptyset(\mathcal{R}_{sd})$ to satisfy the two boundary conditions in Definition 8. Since the link and route resistances decrease as $\mathcal{C}$ increases in size, the route vulnerability increases continuously, satisfying the third condition of Definition 8.

For independent path routing, the subgraph G_{sd} can be decomposed into individual single-path routes. For each path $\pi \in \mathcal{R}_{sd}$, the circuit mapping using the planar dual in Section 4.2.1 can be applied to the single-path route $\{\pi\}$. As illustrated in Fig. 3, the equivalent resistance $R_C(\{\pi\})$ of the single-path route $\{\pi\}$ is equal to that of a parallel combination of link resistors $R_C(i,j)$ plus the series link resistor $R_C(s,d)$ for the end-to-end secure link (s,d), given by

$$R_C(\{\pi\}) = R_C(s,d) + \left(\sum_{(i,j)\in\pi} R_C(i,j)^{-1} \right)^{-1}. \tag{6}$$

As in Section 4.1, the route vulnerability is given by the weighted sum of the vulnerability for the single path routes, yielding

$$V^I_{sd}(\mathcal{C})_{\text{CIR}} = \sum_{\pi\in\mathcal{R}_{sd}} \frac{f_\pi}{R_\emptyset(\{\pi\})} \left(\frac{1+R_\emptyset(\{\pi\})}{1+R_C(\{\pi\})} - 1 \right). \tag{7}$$

For dependent path routing, the entire route subgraph G_{sd} is considered in the evaluation of the route resistance $R_C(\mathcal{R}_{sd})$, yielding

$$V^D_{sd}(\mathcal{C})_{\text{CIR}} = \frac{1}{R_\emptyset(\mathcal{R}_{sd})} \left(\frac{1+R_\emptyset(\mathcal{R}_{sd})}{1+R_C(\mathcal{R}_{sd})} - 1 \right). \tag{8}$$

4.3 Comparison of RVM Realizations

We provide a pair of examples to illustrate cases in which the RVMs in Sections 4.1 and 4.2 are evaluated using the provided definitions. Each of the examples demonstrates a scenario where one metric provides more information to the adversary in choosing which nodes to capture. Both examples are presented using the network illustrated in Fig. 4 for a source s and destination d with intermediate nodes m_1 and m_2. The number of shared keys for each link is illustrated in Fig. 4, and the costs to capture s, m_1, m_2, and d are assumed to be infinity. The other nodes present in the network have unit cost to capture and are of two types, represented by nodes n_1 and n_2, and the adversary must choose between n_1 and n_2 by computing the node values $\nu_{n_1}(\emptyset)$ and $\nu_{n_2}(\emptyset)$ using Definition 9. For each example, the node value is computed using both the set and circuit theoretic RVM realizations.

For the first example, we consider a single two-hop path through an intermediate node m_2, ignoring the path through m_1. Suppose that $|\mathcal{K}_{sm_2} \cap \mathcal{K}_{n_1}| = 2$ and $|\mathcal{K}_{m_2 d} \cap \mathcal{K}_{n_2}| = 1$. The node values $\nu_{n_1}(\emptyset)$ and $\nu_{n_2}(\emptyset)$ are computed using the set theoretic RVM given by (4) and the

TABLE 2
The Route Vulnerability is Compared for the Two Examples Using the Network in Fig. 4 Using the Set and Circuit Theoretic RVM Realizations

	First Example		Second Example	
Node Value	Set, (4)	Circuit, (7)	Set, (5)	Circuit, (8)
$\nu_{n_1}(\emptyset)$	1/2	1/8	3/8	1/26
$\nu_{n_2}(\emptyset)$	1/2	2/9	7/32	1/26

circuit theoretic RVM given by (7), and the results are given in the second and third columns of Table 2.

For the second example, we consider two dependent paths π_1 and π_2 traversing intermediate nodes m_1 and m_2, respectively. Suppose that $|\mathcal{K}_{sm_2} \cap \mathcal{K}_{n_1}| = 3$, $|\mathcal{K}_{sm_1} \cap \mathcal{K}_{n_2}| = 2$, $|\mathcal{K}_{m_1 d} \cap \mathcal{K}_{n_2}| = 1$, and the remaining intersection sets are empty. The node values $\nu_{n_1}(\emptyset)$ and $\nu_{n_2}(\emptyset)$ are computed using the set theoretic RVM given by (5) and the circuit theoretic RVM given by (8), and the results are given in the fourth and fifth columns of Table 2.

The two examples using the network in Fig. 4 and the corresponding node values in Table 2 illustrate that each of the set and circuit theoretic RVMs can lead to ties in node value. If the use of one metric yields a tie, the other metric can be evaluated as a tiebreaker. Hence, the two metrics are complementary in the evaluation of route vulnerability.

5 Uncertainty in RVM Parameters Due to Privacy-Preserving Set Intersection

In order for an adversary to mount a node capture attack using the GNAVE algorithm, the contribution $V_{sd}(\mathcal{C} \cup \{n\}) - V_{sd}(\mathcal{C})$ to the incremental node value $\nu_n(\mathcal{C})$ of node $n \in \mathcal{N}$ must be computed using Definition 9 with an RVM realization that satisfies Definition 8. Both the set and circuit theoretic RVM realizations in Section 4 require the adversary to compute the quantities $|\mathcal{K}_{ij}|$ and $|\mathcal{K}_{ij} \cap \mathcal{K}_{\mathcal{C}}|$ for each link (i, j) in the route $\mathcal{R}_{sd}$. As proven in [4], the set $\mathcal{K}_i \cap \mathcal{K}_{\mathcal{C}}$ can be computed for any $i \in \mathcal{N}$ by the adversary with captured nodes $\mathcal{C}$. Hence, the quantity $|\mathcal{K}_{ij} \cap \mathcal{K}_{\mathcal{C}}|$ can always be computed using the equality

$$\mathcal{K}_{ij} \cap \mathcal{K}_{\mathcal{C}} = (\mathcal{K}_i \cap \mathcal{K}_{\mathcal{C}}) \cap (\mathcal{K}_j \cap \mathcal{K}_{\mathcal{C}}).$$

However, if the network nodes in $\mathcal{N}$ are using a privacy-preserving set intersection protocol according to Definition 4, the quantity $|\mathcal{K}_{ij}|$ cannot be computed deterministically. We thus demonstrate how this required quantity can be estimated probabilistically. In what follows, we assume that each set $\mathcal{K}_i$ is randomly and independently selected from $\mathcal{K}$ as in [3] and that the quantities $k_i = |\mathcal{K}_i|$ and $k = |\mathcal{K}|$ are publicly known.

A probabilistic estimate $\hat{k}_{ij}$ of the quantity $|\mathcal{K}_{ij}|$ can be computed using the probability distribution $\Pr[|\mathcal{K}_{ij}| = k_{ij}]$ using the known parameters $k_{i\mathcal{C}} = |\mathcal{K}_i \cap \mathcal{K}_{\mathcal{C}}|$, $k_{j\mathcal{C}} = |\mathcal{K}_j \cap \mathcal{K}_{\mathcal{C}}|$, and $k_{ij\mathcal{C}} = |\mathcal{K}_{ij} \cap \mathcal{K}_{\mathcal{C}}|$. This probability is exactly the probability that $(k_{ij} - k_{ij\mathcal{C}})$ of the $(k_i - k_{i\mathcal{C}})$ keys in $\mathcal{K}_i$ not known to the adversary are equal to $(k_{ij} - k_{ij\mathcal{C}})$ of the $(k_j - k_{j\mathcal{C}})$ keys in $\mathcal{K}_j$ not known to the adversary. Letting $k_{\mathcal{C}} = |\mathcal{K}_{\mathcal{C}}|$, the desired probability can be computed as

$$\Pr\left[|\mathcal{K}_{ij}| = k_{ij}\right] = \binom{k_j - k_{j\mathcal{C}}}{k_{ij} - k_{ij\mathcal{C}}} \left(\frac{k_i - k_{i\mathcal{C}}}{k - k_{\mathcal{C}}}\right)^{k_{ij} - k_{ij\mathcal{C}}} \times \left(1 - \frac{k_i - k_{i\mathcal{C}}}{k - k_{\mathcal{C}}}\right)^{k_j - k_{j\mathcal{C}} - k_{ij} + k_{ij\mathcal{C}}} \tag{9}$$

for $k_{ij} = k_{ij\mathcal{C}}, \ldots, k_j - k_{j\mathcal{C}} + k_{ij\mathcal{C}}$.

We compute the estimate $\hat{k}_{ij}$ as the expected value of $|\mathcal{K}_{ij}|$, conditioned on the fact that $|\mathcal{K}_{ij}| > k_{ij\mathcal{C}}$ since $|\mathcal{K}_{ij}|$ is only unknown if $k_j > k_{j\mathcal{C}}$. The estimate $\hat{k}_{ij}$ is thus computed as the expected value of the random variable with probability distribution $\Pr[|\mathcal{K}_{ij}| = k_{ij}] / \Pr[|\mathcal{K}_{ij}| > k_{ij\mathcal{C}}]$, subject to $|\mathcal{K}_{ij}| > k_{ij\mathcal{C}}$, using (9), yielding

$$\hat{k}_{ij} = k_{ij\mathcal{C}} + \frac{(k_i - k_{i\mathcal{C}})(k_j - k_{j\mathcal{C}})}{(k - k_{\mathcal{C}})\left(1 - \left(1 - \frac{k_i - k_{i\mathcal{C}}}{k - k_{\mathcal{C}}}\right)^{k_j - k_{j\mathcal{C}}}\right)}. \tag{10}$$

The estimate $\hat{k}_{ij}$ of $|\mathcal{K}_{ij}|$ using (10) can then be used to estimate the route vulnerability $V_{sd}(\mathcal{C})$. However, in order to estimate the incremental node value $\nu_n(\mathcal{C})$ of each node $n \in \mathcal{N} \setminus \mathcal{C}$, the route vulnerability $V_{sd}(\mathcal{C} \cup \{n\})$ must also be estimated, where the union $\mathcal{K}_{\mathcal{C} \cup \{n\}}$ cannot be computed deterministically.

We note that for any i, j, $n \in \mathcal{N}$, the number of keys securing the link (i, j) if n is added to $\mathcal{C}$ is given by

$$\begin{aligned} |\mathcal{K}_{ij} \setminus \mathcal{K}_{\mathcal{C} \cup \{n\}}| &= |\mathcal{K}_{ij}| - |\mathcal{K}_{ij} \cap (\mathcal{K}_{\mathcal{C}} \cup \mathcal{K}_n)| \\ &= |\mathcal{K}_{ij}| - |\mathcal{K}_{ij} \cap \mathcal{K}_{\mathcal{C}}| \\ &\quad - |\mathcal{K}_{ij} \cap \mathcal{K}_n| + |\mathcal{K}_{ij} \cap \mathcal{K}_n \cap \mathcal{K}_{\mathcal{C}}|. \end{aligned} \tag{11}$$

Since the quantities $k_{ij\mathcal{C}} = |\mathcal{K}_{ij} \cap \mathcal{K}_{\mathcal{C}}|$ and $k_{ijn\mathcal{C}} = |\mathcal{K}_{ij} \cap \mathcal{K}_n \cap \mathcal{K}_{\mathcal{C}}|$ are known, and an estimate $\hat{k}_{ij}$ of $|\mathcal{K}_{ij}|$ has already been computed in (10), the remaining quantity to estimate is $|\mathcal{K}_{ij} \cap \mathcal{K}_n|$. We let $Q(k_{ijn})$ denote the probability $\Pr[|\mathcal{K}_{ij} \cap \mathcal{K}_n| = k_{ijn}]$ and $Q(k_{ijn}|k_{ij})$ denote the similar probability conditioned on the event that $|\mathcal{K}_{ij}| = k_{ij}$, computed as

$$Q(k_{ijn}) = \sum_{k_{ij} = k_{ij\mathcal{C}}}^{k_j - k_{j\mathcal{C}} + k_{ij\mathcal{C}}} Q(k_{ijn}|k_{ij}) \Pr[|\mathcal{K}_{ij}| = k_{ij}]. \tag{12}$$

Similar to the computation in (9), the conditional probability $Q(k_{ijn}|k_{ij})$ is computed as

$$Q(k_{ijn}|k_{ij}) = \binom{k_n - k_{n\mathcal{C}}}{k_{ijn}} \left(\frac{k_{ij} - k_{ij\mathcal{C}}}{k - k_{\mathcal{C}}}\right)^{k_{ijn}} \times \left(1 - \frac{k_{ij} - k_{ij\mathcal{C}}}{k - k_{\mathcal{C}}}\right)^{k_n - k_{n\mathcal{C}} - k_{ijn}}. \tag{13}$$

An estimate $\hat{k}_{ijn}$ of $|\mathcal{K}_{ij} \cap \mathcal{K}_n|$ is computed conditioned on the event that $|\mathcal{K}_{ij}| > k_{ij\mathcal{C}}$ as before. The estimate $\hat{k}_{ijn}$ is thus computed as the expected value of the random variable

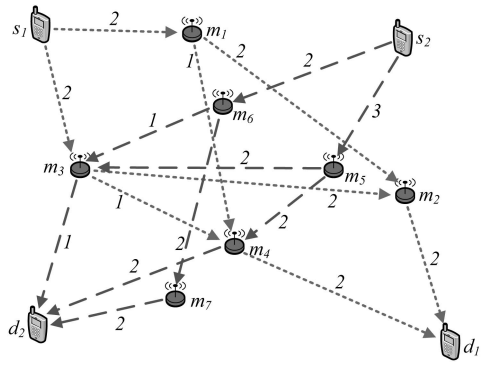

Fig. 5. Two sources s_1 and s_2 send messages to destinations d_1 and d_2, respectively, using an independent path routing protocols based on geographic forwarding to multiple next-hop neighbors. Each link (i, j) is labeled with the number of shared keys $|\mathcal{K}_{ij}|$. The end-to-end secure links, not illustrated, have $|\mathcal{K}^E_{s_1d_1}| = 2$ and $|\mathcal{K}^E_{s_2d_2}| = 2$ shared keys each.

with probability distribution $Q(k_{ijn})/\Pr[|\mathcal{K}_{ij}| > k_{ij\mathcal{C}}]$, subject to $k_{ij} > k_{ij\mathcal{C}}$, using (9), (12), and (13) yielding

$$\begin{aligned}\hat{k}_{ijn} &= \sum_{k_{ijn}=0}^{k_n-k_{n\mathcal{C}}} \frac{k_{ijn} \sum_{k_{ij}=k_{ij\mathcal{C}}+1}^{k_j-k_{j\mathcal{C}}+k_{ij\mathcal{C}}} Q(k_{ijn}|k_{ij}) \Pr[|\mathcal{K}_{ij}| = k_{ij}]}{\Pr[|\mathcal{K}_{ij}| > k_{ij\mathcal{C}}]} \\ &= \sum_{k_{ij}=k_{ij\mathcal{C}}+1}^{k_j-k_{j\mathcal{C}}+k_{ij\mathcal{C}}} \frac{\Pr[|\mathcal{K}_{ij}| = k_{ij}]}{\Pr[|\mathcal{K}_{ij}| > k_{ij\mathcal{C}}]} \sum_{k_{ijn}=0}^{k_n-k_{n\mathcal{C}}} k_{ijn} Q(k_{ijn}|k_{ij}) \\ &= \frac{k_n - k_{n\mathcal{C}}}{k - k_{\mathcal{C}}} \sum_{k_{ij}=k_{ij\mathcal{C}}+1}^{k_j-k_{j\mathcal{C}}+k_{ij\mathcal{C}}} \frac{\Pr[|\mathcal{K}_{ij}| = k_{ij}](k_{ij} - k_{ij\mathcal{C}})}{\Pr[|\mathcal{K}_{ij}| > k_{ij\mathcal{C}}]} \\ &= \frac{(k_n - k_{n\mathcal{C}})(\hat{k}_{ij} - k_{ij\mathcal{C}})}{k - k_{\mathcal{C}}},\end{aligned} \tag{14}$$

where $\hat{k}_{ij}$ is the estimate given in (10).

The estimates $\hat{k}_{ij}$ and $\hat{k}_{ijn}$ are then used to estimate the incremental node value $\nu_n(\mathcal{C})$ of each node $n \in \mathcal{N} \setminus \mathcal{C}$ using Definition 9 with the corresponding route vulnerability definitions in Section 4. We note that the contribution of a node toward the compromise of a link, path, or route is deterministic if the captured node is incident to the link, path, or route of interest. Hence, at early stages of the attack, it is likely that captured nodes will be located along paths from source nodes to destination nodes. The adversary will, however, learn significantly more information about the remainder of the network by capturing one node at a time using the GNAVE algorithm with the vulnerability estimates obtained herein.

6 Examples and Simulation Study

In this section, we illustrate the application of the RVM $V_{sd}(\mathcal{C})$ and the GNAVE algorithm. We first present two small-scale examples using independent and dependent path routing and the set and circuit theoretic RVMs. We then provide simulation results to illustrate the effect of node capture attacks in a large-scale wireless network under various different key assignment and routing models.

TABLE 3
Route Vulnerabilities and Node Values are Computed for the Set and Circuit Theoretic RVMs for the Network in Fig. 5, Rounding Each Quantity to the Nearest 0.001

i	$V^I_{sd}(\{i\})_{\text{SET}}$		$\nu_i(\emptyset)_{\text{SET}}$	$V^I_{sd}(\{i\})_{\text{CIR}}$		$\nu_i(\emptyset)_{\text{CIR}}$
	s_1, d_1	s_2, d_2		s_1, d_1	s_2, d_2	
s_1	1.000	0.100	1.100	1.000	0.004	1.004
s_2	0.400	1.000	1.400	0.024	1.000	1.024
m_1	0.700	0.100	0.800	0.243	0.004	0.247
m_2	0.950	0.500	1.450	0.913	0.025	0.937
m_3	0.600	0.783	1.383	0.237	0.248	0.485
m_4	0.775	0.975	**1.750**	0.247	0.914	**1.161**
m_5	0.300	0.800	1.100	0.021	0.272	0.293
m_6	0.300	0.767	1.067	0.010	0.243	0.254
m_7	0.500	0.500	1.000	0.019	0.207	0.226

6.1 Example: Multipath Geographic Forwarding

We illustrate a node capture attack using the GNAVE algorithm with the set and circuit theoretic RVMs presented in Sections 4.1 and 4.2, respectively, for a wireless network using multipath geographic forwarding. In this example, we construct independent path routes using a multipath geographic forwarding algorithm in which each node forwards the corresponding message to the two next-hop neighbors nearest the destination node, similar to the idea in GBR [13]. We consider the network topology given in Fig. 5 with source-destination routing pairs $\mathcal{T} = \{(s_1, d_1), (s_2, d_2)\}$. The additional end-to-end security mechanism discussed in Section 2.2 is used by each source-destination pair, and keys are assigned to nodes in the network as follows:

$\mathcal{K}_{s_1} = \{k_2, k_7, k_8, k_{10}\}$	$\mathcal{K}_{s_2} = \{k_3, k_4, k_6, k_{11}\}$
$\mathcal{K}_{d_1} = \{k_1, k_2, k_{10}, k_{11}\}$	$\mathcal{K}_{d_2} = \{k_3, k_8, k_9, k_{11}\}$
$\mathcal{K}_{m_1} = \{k_8, k_9, k_{10}, k_{12}\}$	$\mathcal{K}_{m_2} = \{k_2, k_6, k_9, k_{10}\}$
$\mathcal{K}_{m_3} = \{k_2, k_3, k_6, k_7\}$	$\mathcal{K}_{m_4} = \{k_3, k_5, k_{10}, k_{11}\}$
$\mathcal{K}_{m_5} = \{k_3, k_4, k_5, k_6\}$	$\mathcal{K}_{m_6} = \{k_1, k_6, k_{11}, k_{12}\}$
$\mathcal{K}_{m_7} = \{k_1, k_7, k_8, k_{11}\}$	

To illustrate the security of each link using the assigned keys above, we note that nodes s_1 and m_1 share keys $\mathcal{K}_{s_1m_1} = \{k_8, k_{10}\}$, so the link (s_1, m_1) is secure as long as $\{k_8, k_{10}\} \not\subseteq \mathcal{K}_{\mathcal{C}}$.

Assuming the messages traversing different paths through the network are independently secured, the route vulnerability of the two routes $\mathcal{R}_{s_1d_1}$ and $\mathcal{R}_{s_2d_2}$ can be computed using (4) or (7) by individually considering the four paths and the end-to-end secure link in each route. The route vulnerabilities $V^I_{sd}(\mathcal{C})_{\text{SET}}$ and $V^I_{sd}(\mathcal{C})_{\text{CIR}}$ and the corresponding node values $\nu_i(\mathcal{C})_{\text{SET}}$ and $\nu_i(\mathcal{C})_{\text{CIR}}$ computed using Definition 9 are provided in Table 3. In computing the node value and considering which nodes can appear in $\mathcal{C}$, we assume that the node capture cost w_i for each source s_j and intermediate node m_j is unity, while that of each destination node is infinity.

To demonstrate the computation of quantities in Table 3, we consider the source-destination pair (s_1, d_1) in the second column and compute the route vulnerability

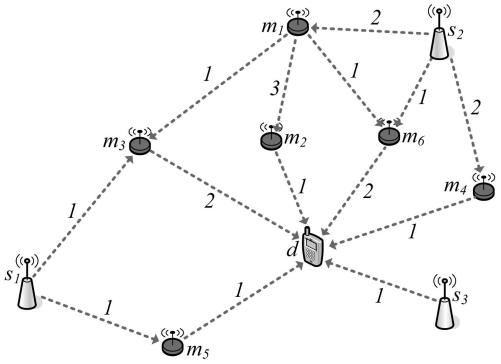

Fig. 6. A destination node d receives messages from source nodes s_1, s_2, and s_3, with copies of the same data, using randomized network coding. Each link (i, j) is labeled with the number of shared keys $|\mathcal{K}_{ij}|$.

resulting from the capture of node m_4. The route $\mathcal{R}_{s_1 d_1}$ consists of four independent paths:

$$\pi_1 = \{(s_1, m_1), (m_1, m_2), (m_2, d_1)\},$$
$$\pi_2 = \{(s_1, m_1), (m_1, m_4), (m_4, d_1)\},$$
$$\pi_3 = \{(s_1, m_3), (m_3, m_2), (m_2, d_1)\},$$
$$\pi_4 = \{(s_1, m_3), (m_3, m_4), (m_4, d_1)\},$$

each corresponding to an independent single-path route. We assume that $f_{\pi_i} = f_{s_1 d_1} = 1/4$.

To compute the set theoretic vulnerability $V^I_{s_1 d_1}(\{m_4\})_{\text{SET}}$ using Fig. 5, we first compute $\phi_{s_1 d_1}(\{m_4\})$ as 1/2, the ϕ values for path π_1 as 1/2, 1/2, and 1/2, the ϕ values for path π_2 as 1/2, 1, and 1, the ϕ values for path π_3 as 0, 0, and 1/2, and the ϕ values for path π_4 as 0, 1, and 1, implying that paths π_2 and π_4 are compromised. From (4), the vulnerability is computed as $V^I_{s_1 d_1}(\{m_4\})_{\text{SET}} = 31/40 = 0.775$, as indicated in Table 3.

To compute the circuit theoretic vulnerability $V^I_{s_1 d_1}(\{m_4\})_{\text{CIR}}$, we first compute the initial resistances $R_\emptyset(\{\pi_i\})$ for each of the four paths, noting that π_1 and π_3 each consist of three links with link resistance 2, while π_2 and π_4 each consist of two links with link resistance 2 and one with link resistance 1. Hence, the resistance values of the four paths are computed using (6), including the end-to-end link of resistance $R_\emptyset(s_1, d_1) = 2$, as $R_\emptyset(\{\pi_1\}) = R_\emptyset(\{\pi_3\}) = 8/3$, $R_\emptyset(\{\pi_2\}) = R_\emptyset(\{\pi_4\}) = 5/2$. When $\mathcal{C} = \{m_4\}$, the link resistance values become 1, 1, and 1 for π_1, 1, 0, and 0 for π_2, 2, 2, and 1 for π_3, and 2, 0, and 0 for π_4. The updated resistance values of the four paths are computed using (6), including the updated end-to-end link of resistance $R_{\{m_4\}}(s_1, d_1) = 1$ as $R_{\{m_4\}}(\pi_1) = 4/3$, $R_{\{m_4\}}(\pi_2) = R_{\{m_4\}}(\pi_4) = 1$, and $R_{\{m_4\}}(\pi_3) = 3/2$. From (7), the vulnerability is computed as $V^I_{s_1 d_1}(\{m_4\})_{\text{CIR}} = 277/1{,}120 \approx 0.247$, as indicated in Table 3.

As indicated in Table 3, the first node added to $\mathcal{C}$ using the GNAVE algorithm under both the set and circuit theoretic vulnerability functions is node m_4.

6.2 Example: Distributed Data Access Using Network Coding

We illustrate a node capture attack using the GNAVE algorithm with the set and circuit theoretic RVMs presented in Sections 4.1 and 4.2, respectively, for a network with three sources sending the same set of messages using network coding. In this example, we construct dependent path routes using a randomized network coding algorithm [16] in which each node forwards a different linear combination of previously received messages in the same message batch along each secure link. We consider the network topology given in Fig. 6 with keys assigned to nodes in the network as follows:

$\mathcal{K}_{s_1} = \{k_7, k_8, k_{11}, k_{13}\}$	$\mathcal{K}_{s_2} = \{k_3, k_5, k_9, k_{13}\}$
$\mathcal{K}_{s_3} = \{k_2, k_7, k_8, k_{11}\}$	$\mathcal{K}_d = \{k_1, k_2, k_{12}, k_{14}\}$
$\mathcal{K}_{m_1} = \{k_3, k_4, k_5, k_{10}\}$	$\mathcal{K}_{m_2} = \{k_3, k_4, k_5, k_{12}\}$
$\mathcal{K}_{m_3} = \{k_2, k_4, k_{11}, k_{14}\}$	$\mathcal{K}_{m_4} = \{k_3, k_9, k_{10}, k_{12}\}$
$\mathcal{K}_{m_5} = \{k_5, k_6, k_{12}, k_{13}\}$	$\mathcal{K}_{m_6} = \{k_2, k_4, k_{13}, k_{14}\}$

TABLE 4
Node Values, Equal to the Route Vulnerabilities, are Computed for the Set and Circuit Theoretic RVMs for the Network in Fig. 6, Rounding Each Quantity to the Nearest 0.001

i	$\nu_i(\emptyset)_{\text{SET}}$	$\nu_i(\emptyset)_{\text{CIR}}$	i	$\nu_i(\emptyset)_{\text{SET}}$	$\nu_i(\emptyset)_{\text{CIR}}$
m_1	0.438	0.061	m_4	0.500	0.092
m_2	0.625	0.145	m_5	0.625	0.156
m_3	0.667	0.237	m_6	**0.792**	**0.342**

Since network coding is used to construct each transmitted packet as a function of the entire batch of messages, packets traversing different paths are dependent, even though links are independently secured. Furthermore, since the three sources s_1, s_2, and s_3 act as a single information source, we can treat the message traversal through the network as a single dependent route, effectively joining the source nodes s_1, s_2, and s_3 into a single source s. Hence, the route vulnerability of the route $\mathcal{R}_{sd}$ can be computed using (5) or (8). The route vulnerabilities $V^D_{sd}(\mathcal{C})_{\text{SET}}$ and $V^D_{sd}(\mathcal{C})_{\text{CIR}}$ and the corresponding node values $\nu_i(\mathcal{C})_{\text{SET}}$ and $\nu_i(\mathcal{C})_{\text{CIR}}$ computed using Definition 9 are provided in Table 4. In computing the node value and considering which nodes can appear in $\mathcal{C}$, we assume that the node capture cost w_i for each intermediate node m_j is unity, while that of each source s_j and the destination node d is infinity.

To demonstrate the computation of quantities in Table 4, we evaluate the route vulnerability due to the capture of node m_6, which is the first node added to $\mathcal{C}$ using the GNAVE algorithm under both the set and circuit theoretic vulnerability functions. To compute the set theoretic vulnerability $V^I_{sd}(\{m_6\})_{\text{SET}}$ for the network in Fig. 6, we note that the route $\mathcal{R}_{sd}$ consists of eight paths

$$\pi_1 = \{(s_1, m_3), (m_3, d)\},$$
$$\pi_2 = \{(s_1, m_5), (m_5, d)\},$$
$$\pi_3 = \{(s_2, m_4), (m_4, d)\},$$
$$\pi_4 = \{(s_2, m_6), (m_6, d)\},$$
$$\pi_5 = \{(s_2, m_1), (m_1, m_2), (m_2, d)\},$$
$$\pi_6 = \{(s_2, m_1), (m_1, m_3), (m_3, d)\},$$
$$\pi_7 = \{(s_2, m_1), (m_1, m_6), (m_6, d)\},$$
$$\pi_8 = \{(s_3, d)\},$$

where the end-to-end link (s, d) is already included as the path π_8 joining s_3 to d. By inspection of the collection of paths and the keys assigned to each node, we compute the ϕ values for each path as 0 and 1 for π_1, 1 and 0 for π_2, 0

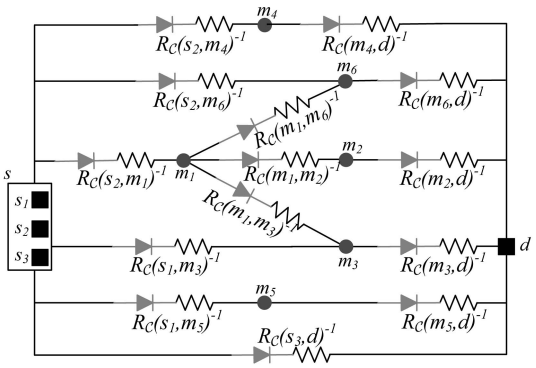

Fig. 7. The dual circuit $\mathcal{E}^*_{sd}$ corresponding to the network in Fig. 6 is illustrated with the three sources s_1, s_2, and s_3 combined into a single source s.

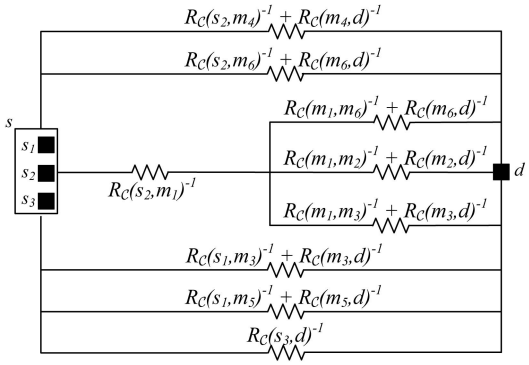

Fig. 8. The circuit $\mathcal{E}^*_{sd}$ in Fig. 7 is simplified by interpreting the effect of the series diodes and eliminating them from the circuit.

and 0 for π_3, 1 and 1 for π_4, 0, 1/3, and 0 for π_5, 0, 1, and 1 for π_6, 0, 1, and 1 for π_7, and 1 for π_8. From (5) with $f_{sd} = 0$, the vulnerability is computed as $V^D_{sd}(\{m_6\})_{\text{SET}} = 19/24 \approx 0.792$, as indicated in Table 4.

To compute the circuit theoretic vulnerability $V^D_{sd}(\{m_6\})_{\text{CIR}}$, we first derive a formula to derive the route resistance $R_{\mathcal{C}}(\mathcal{R}_{sd})$ as a function of the set of link resistance values $R_{\mathcal{C}}(i,j)$. Though the route subgraph G_{sd} in Fig. 6 is a planar graph, we make use of the circuit dual technique discussed in Section 4.2.2 for the purpose of illustration. We first construct the dual circuit $\mathcal{E}^*_{sd}$ by replacing each edge in G_{sd} by a series resistor-diode pair where the resistance is equal to $R_{\mathcal{C}}(i,j)^{-1}$. The circuit is first simplified by combining the three sources s_1, s_2, and s_3 into a single source s and rearranging the nodes and edges for clarity as in Fig. 7. The circuit is then simplified further by interpreting the effect of the diodes in series with the link resistors, as illustrated in Fig. 8. The simplification is described as follows:

In order to determine the effect of the diodes in the circuit $\mathcal{E}^*_{sd}$, we consider the current flowing along each path. The paths π_2, π_3, and π_8 are disjoint from the remaining paths, so the diodes have no effect and can be removed, yielding paths of resistance $(R_{\mathcal{C}}(s_1,m_5)^{-1} + R_{\mathcal{C}}(m_5,d)^{-1})$, $(R_{\mathcal{C}}(s_2,m_4)^{-1} + R_{\mathcal{C}}(m_4,d)^{-1})$, and $R_{\mathcal{C}}(s_3,d)^{-1}$, respectively. A current flowing along the path π_1 from node s_1 reaches node m_3, seeing an infinite resistance in the direction of node m_1 and a finite resistance to the destination d. The path π_1 can thus be separated as a disjoint path from s to d with resistance $(R_{\mathcal{C}}(s_1,m_1)^{-1} + R_{\mathcal{C}}(m_1,d)^{-1})$. The path π_4 can similarly be replaced by a disjoint path with resistance $(R_{\mathcal{C}}(s_2,m_6)^{-1} + R_{\mathcal{C}}(m_6,d)^{-1})$. A current flowing from s toward node m_1 is not influenced by the diode in series with the resistance $R_{\mathcal{C}}(s_2,m_1)^{-1}$, so the diode can be removed. From node m_1, there are three disjoint paths to d, and the diodes in the remaining paths can similarly be eliminated. The series resistances of each of the three disjoint paths from m_1 to d can then each be simplified, yielding three parallel resistors of resistance $(R_{\mathcal{C}}(m_1,m_2)^{-1} + R_{\mathcal{C}}(m_2,d)^{-1})$, $(R_{\mathcal{C}}(m_1,m_2)^{-1} + R_{\mathcal{C}}(m_2,d)^{-1})$, and $(R_{\mathcal{C}}(m_1,m_6)^{-1} + R_{\mathcal{C}}(m_6,d)^{-1})$. The resistive circuit resulting from these simplifications is illustrated in Fig. 8. The equivalent resistance $R_{\mathcal{C}}(\mathcal{R}_{sd})^{-1}$ of the dual circuit $\mathcal{E}^*_{sd}$ can thus be computed using standard equivalence techniques for series and parallel resistance.

Substituting the corresponding inverse link resistance values into the circuit in Fig. 8 and computing the inverse of the equivalent resistance thus yields the route resistances $R_{\emptyset}(\mathcal{R}_{sd}) = 443/98$ and $R_{\{m_6\}}(\mathcal{R}_{sd}) = 7/6$. The route vulnerability is then computed using (8) as $V^D_{sd}(\{m_5\})_{\text{CIR}} = 1{,}143/3{,}338 \approx 0.342$, as indicated in Table 4.

6.3 Simulation Study: Wireless Sensor Network

We provide simulation results to illustrate a node capture attack using the GNAVE algorithm using the circuit theoretic RVM presented in Section 4.2. A similar study can be performed for the set theoretic RVM presented in Section 4.1. We compare the performance of the attack to node capture attacks using existing node selection metrics.

The simulation was performed for a wireless sensor network of $|\mathcal{N}| = 500$ sensor nodes deployed randomly over a square region with density to yield an average of 25 neighbors per sensor node. Each node $i \in \mathcal{N}$ was randomly assigned a set of $|\mathcal{K}_i| = 50$ keys using key predistribution as in [3]. A subset of $|\mathcal{S}| = 100$ nodes was randomly selected as the set of source nodes, and a subset of $|\mathcal{D}| = 10$ nodes was randomly selected as the set of destination nodes. For each source $s \in \mathcal{S}$, the three nearest destination nodes in $\mathcal{D}$ were chosen as route pairs $(s,d) \in \mathcal{T}$. Each route $\mathcal{R}_{sd}$ was constructed using geographic forwarding with a hop-count mechanism to avoid routing loops and geographic dead ends due to holes [14]. For both independent and dependent path routing, each node chose three next-hop neighbors closest to the destination and with a lower or equal hop count. For dependent path routing, we assume that any compromised edge cut is sufficient to compromise the route.

We simulated the node capture attacks using multiple strategies for both independent and dependent path routing. We simulated secure link establishment using public label exchange without end-to-end security, public label exchange with end-to-end security, and privacy-preserving set intersection without end-to-end security using the estimation techniques in Section 5. Node capture attacks on each case were simulated for the following five node capture strategies:

1. Nodes are captured independently at random, serving as the baseline performance for the adversary.
2. Nodes are captured iteratively to maximize the number of compromised keys $|\mathcal{K}_{\mathcal{C}}|$ by choosing the

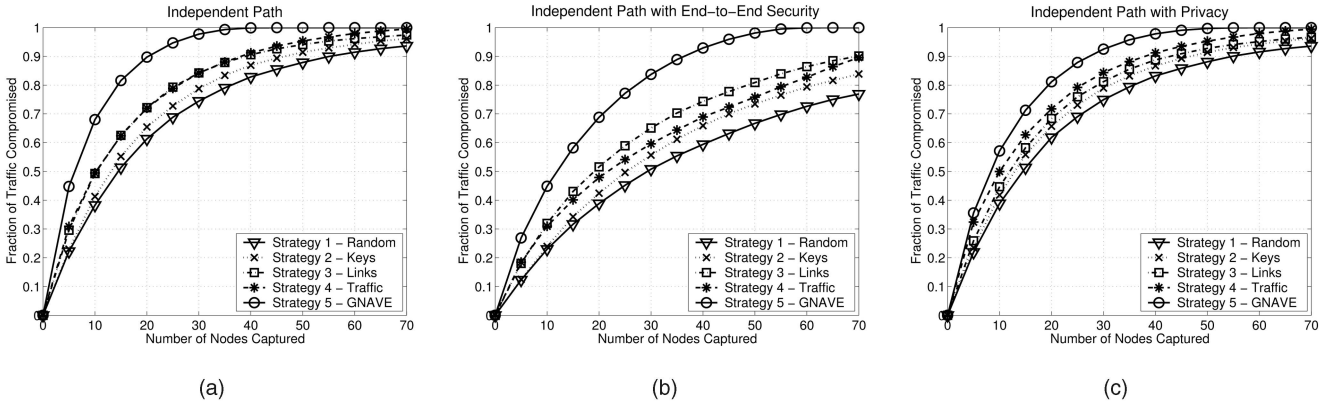

Fig. 9. Node capture attacks using the five strategies are illustrated for a wireless sensor network of $|\mathcal{N}| = 500$ nodes for independent path routing (a) without end-to-end security, (b) with end-to-end security, and (c) using a privacy-preserving set intersection protocol.

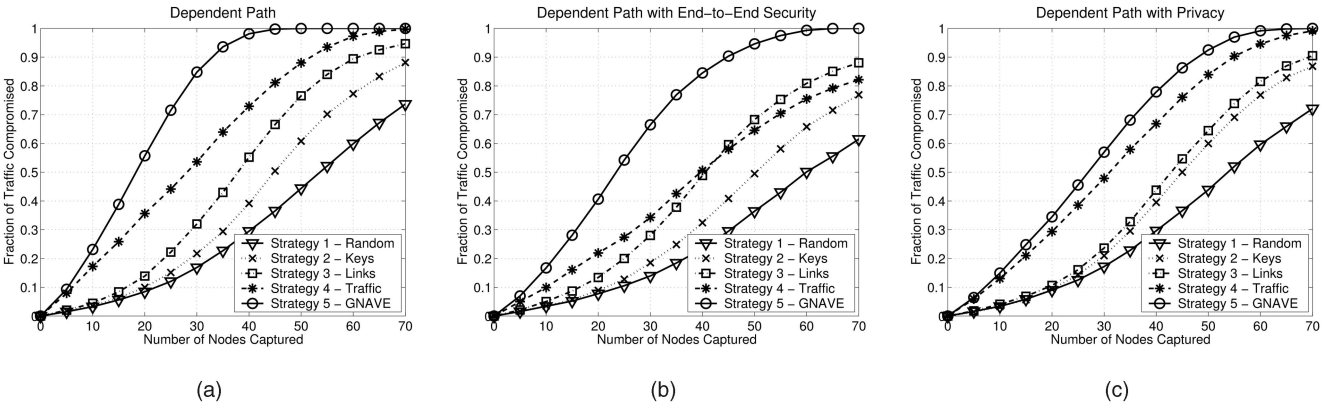

Fig. 10. Node capture attacks using the five strategies are illustrated for a wireless sensor network of $|\mathcal{N}| = 500$ nodes for dependent path routing (a) without end-to-end security, (b) with end-to-end security, and (c) using a privacy-preserving set intersection protocol.

node i with maximum $|\mathcal{K}_i \setminus \mathcal{K}_{\mathcal{C}}|$ at each iteration, independent of the routing protocol. We note that such an attack can be performed deterministically under privacy-preserving protocols [4].

3. Nodes are captured iteratively to maximize the number of compromised links $|L_{\mathcal{C}}|$ by choosing the node i which compromises the maximum number of additional links, independent of the routing protocol. Under privacy-preserving protocols, this attack uses the estimation techniques in Section 5.
4. Nodes are captured iteratively to maximize the amount of network traffic routed through captured nodes, independent of the key assignment protocol.
5. Nodes are captured using the GNAVE algorithm and the RVM $V_{sd}^{I}(\mathcal{C})_{\text{CIR}}$ or $V_{sd}^{D}(\mathcal{C})_{\text{CIR}}$, using information from both the routing and key assignment protocols.

Figs. 9 and 10 illustrate the node capture attacks on independent and dependent path routing, respectively. In each figure, we notice that the node capture attack using the GNAVE algorithm outperforms the remaining attacks. The inclusion of the end-to-end shared keys $\mathcal{K}_{sd}$ in Figs. 9b and 10b show a consistent decrease in the attack performance for all attacks and all routing protocols due to the additional secure end-to-end link that must be compromised in each route. The addition of privacy-preserving set intersection protocols in Figs. 9c and 10c illustrate the increased uncertainty in route vulnerability which slightly degrades the performance of the attack using the GNAVE algorithm and the circuit theoretic vulnerability metric. In comparing Figs. 9 and 10, we notice that the dependence of messages traversing different paths displays a threshold behavior, reducing the vulnerability of routes for small $|\mathcal{C}|$, but only slightly increasing the number of captured nodes $|\mathcal{C}|$ required to compromise all traffic.

7 Conclusion

In this article, we investigated the problem of developing new vulnerability metrics that improve the efficiency of node capture attacks when the routing and key assignment protocols used in a wireless network are jointly analyzed. We proposed a class of route vulnerability metrics (RVMs) to evaluate the effect of node capture attacks on secure network traffic and developed two RVM realizations using set and circuit theoretic interpretations of the compromise of secure network traffic. We formulated the optimal node capture attack using RVM evaluation as a nonlinear integer programming minimization problem and presented the GNAVE algorithm using a greedy heuristic to approximate the NP-hard problem. We demonstrated a probabilistic approach to estimate the route vulnerability when privacy-preserving set intersection protocols are used to hide information from the adversary. Finally, we illustrated node capture attacks using the GNAVE algorithm and compared the performance of the GNAVE algorithm with previously proposed node capture strategies. We provided simulation results to demonstrate the performance gains in using the circuit theoretic RVM, noting that similar results

not included in this article have been obtained using the set theoretic RVM. In the future, the node capture attack framework proposed in this article will assist in the joint design of key assignment and routing protocols for wireless networks that are robust to node capture attacks.

ACKNOWLEDGMENTS

This work was supported in part by the following grants: ONR YIP, N00014-04-1-0479; ARO PECASE, W911NF-05-1-0491; ARL CTA, DAAD19-01-2-001; and ARO MURI, W911NF-07-1-0287. This document was prepared through collaborative participation in the Communications and Networks Consortium sponsored by the US Army Research Laboratory under the Collaborative Technology Alliance Program, DAAD19-01-2-0011. The US Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation thereon. The views and conclusions contained in this document are those of the author and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the US Government. A preliminary version of this material appeared at the 27th IEEE International Conference on Computer Communications (INFOCOM '08) [1].

REFERENCES

[1] P. Tague, D. Slater, J. Rogers, and R. Poovendran, "Vulnerability of Network Traffic Under Node Capture Attacks Using Circuit Theoretic Analysis," *Proc. IEEE INFOCOM '08,* pp. 664-672, Apr. 2008.

[2] A.J. Menezes, P.C. van Oorschot, and S.A. Vanstone, *Handbook of Applied Cryptography.* CRC, 1996.

[3] L. Eschenauer and V.D. Gligor, "A Key-Management Scheme for Distributed Sensor Networks," *Proc. Ninth ACM Conf. Computer and Comm. Security (CCS '02),* pp. 41-47, Nov. 2002.

[4] P. Tague and R. Poovendran, "Modeling Adaptive Node Capture Attacks in Multi-Hop Wireless Networks," *Ad Hoc Networks,* vol. 5, no. 6, pp. 801-814, Aug. 2007.

[5] H. Chan, A. Perrig, and D. Song, "Random Key Predistribution Schemes for Sensor Networks," *Proc. IEEE Symp. Security and Privacy (SP '03),* pp. 197-213, May 2003.

[6] W. Du, J. Deng, Y.S. Han, P.K. Varshney, J. Katz, and A. Khalili, "A Pairwise Key Predistribution Scheme for Wireless Sensor Networks," *ACM Trans. Information and System Security,* vol. 8, no. 2, pp. 228-258, May 2005.

[7] D. Liu, P. Ning, and R. Li, "Establishing Pairwise Keys in Distributed Sensor Networks," *ACM Trans. Information and System Security,* vol. 8, no. 1, pp. 41-77, Feb. 2005.

[8] N. Cai and R.W. Yeung, "Secure Network Coding," *Proc. IEEE Int'l Symp. Information Theory (ISIT '02),* p. 323, June/July 2002.

[9] K. Jain, "Security Based on Network Topology against the Wiretapping Attack," *IEEE Wireless Comm.,* vol. 11, no. 1, pp. 68-71, Feb. 2004.

[10] T.H. Cormen, C.E. Leiserson, and R.L. Rivest, *Introduction to Algorithms.* MIT Press, McGraw-Hill, 2000.

[11] E.M. Royer and C.E. Perkins, "Ad Hoc On-Demand Distance Vector Routing," *Proc. Second IEEE Workshop Mobile Computing Systems and Applications (WMCSA '99),* pp. 90-100, Feb. 1999.

[12] D.B. Johnson, D.A. Maltz, and J. Broch, *DSR: The Dynamic Source Routing Protocol for Multihop Wireless Ad Hoc Networks.* Addison-Wesley, ch. 5, pp. 139-172, 2001.

[13] C. Schurgers and M.B. Srivastava, "Energy Efficient Routing in Wireless Sensor Networks," *Proc. Military Comm. Conf. (MILCOM '01),* pp. 357-361, Oct. 2001.

[14] Y. Yu, R. Govindan, and D. Estrin, "Geographical and Energy Aware Routing: A Recursive Data Dissemination Protocol for Wireless Sensor Networks," Dept. Computer Science, Univ. of California, Los Angeles, Technical Report UCLA/CSD-TR-01-0023, May 2001.

[15] A. Shamir, "How to Share a Secret," *Comm. ACM,* vol. 22, no. 11, pp. 612-613, Nov. 1979.

[16] T. Ho, R. Koetter, M. Medard, D.R. Karger, and M. Effros, "The Benefits of Coding over Routing in a Randomized Setting," *Proc. IEEE Int'l Symp. Information Theory (ISIT '03),* p. 441, June/July 2003.

[17] M. Ramkumar and N. Memon, "An Efficient Random Key Pre-Distribution Scheme," *Proc. IEEE Conf. Global Comm. (GLOBECOM '04),* pp. 2218-2223, Nov./Dec. 2004.

[18] G. Danezis and R. Clayton, "Introducing Traffic Analysis," *Digital Privacy: Theory, Technologies, and Practices,* A. Acquisti, S. Gritzalis, C. Lambrinoudakis, and S. di Vimercati, eds., Auerbach, Dec. 2007.

[19] G. Dobson, "Worst-Case Analysis of Greedy Heuristics for Integer Programming with Nonnegative Data," *Math. of Operations Research,* vol. 7, no. 4, pp. 515-531, Nov. 1982.

[20] V. Chvatel, "Greedy Heuristic for the Set-Covering Problem," *Math. of Operations Research,* vol. 4, no. 3, pp. 233-235, Aug. 1979.

[21] R. Diestel, *Graph Theory,* third ed. Springer, 2005.

[22] D.F. Tuttle Jr., *Electric Networks: Analysis and Synthesis.* McGraw-Hill, 1965.

Patrick Tague received the BS degrees in mathematics and computer engineering from the University of Minnesota, Twin Cities, in 2003 and the MS degree from the University of Washington, Seattle, in 2007. He is a PhD candidate in the Department of Electrical Engineering, University of Washington. His current research interests include analytical modeling of practical key distribution systems for wireless ad hoc and sensor networks and attacks and defense mechanisms for distributed wireless networks. He is a student member of the IEEE.

David Slater received the BS degree from the University of Washington, Seattle, in 2006. He is a PhD candidate in the Department of Electrical Engineering, University of Washington. His current research interests include modeling and analysis of vulnerability and defense mechanisms for wireless ad hoc and sensor networks. He is a student member of the IEEE.

Jason Rogers is a researcher with the Center for High Assurance Computer Systems, Information Technology Division, US Naval Research Laboratory, Washington, District of Columbia.

Radha Poovendran received the PhD degree in electrical engineering from the University of Maryland, College Park, in 1999. He is an associate professor and the founding director of the Network Security Lab (NSL), Department of Electrical Engineering, University of Washington, Seattle. His research interests include the areas of applied cryptography for multiuser environment, wireless networking, and applications of information theory to security. He is a coeditor of the book *Secure Localization and Time Synchronization in Wireless Ad Hoc and Sensor Networks* (Springer-Verlag, 2007). He was a recipient of the NSA Rising Star Award and Faculty Early Career Awards, including the National Science Foundation CAREER Award in 2001, the Army Research Office YIP Award in 2002, the Office of Naval Research YIP Award in 2004, PECASE in 2005 for his research contributions to multiuser security, and a Graduate Mentor Recognition Award from the University of California, San Diego in 2006. He cochaired the first ACM Conference on Wireless Network Security (WiSec) in 2008. He is a senior member of the IEEE and the IEEE Computer Society.