On the Effect of Node Misbehavior in Ad Hoc Networks

Matthias Hollick, Jens Schmitt, Christian Seipl, and Ralf Steinmetz
Multimedia Communications Lab (KOM), Department of Electrical Engineering and Information Technology
Darmstadt University of Technology, Merckstrasse 25, 64283 Darmstadt, Germany
http://www.kom.tu-darmstadt.de
{Matthias.Hollick, Jens.Schmitt, Ralf.Steinmetz}@KOM.tu-darmstadt.de

Abstract -- **The dependability of the routing system in ad hoc networks inherently relies on node behavior. In order to support multi-hop operation in the network, most ad hoc routing algorithms assume well-behaving nodes. However, in reality there may exist constrained, selfish or malicious nodes. We discuss the influence of node misbehavior on the routing process. In particular, we derive a classification for misbehaving nodes and extend an analytical model of the route acquisition process executed by the Ad hoc On-demand Distance Vector (AODV) routing protocol to cover different classes of misbehavior. The validation of the behavior model, and the clarification of the impact misbehaving nodes impose onto the routing process, is completed using an experimental analysis.**

Keywords -- **Ad Hoc Routing, Dependability, Node Misbehavior, Model Development, Model Validation, Experimentation.**

I. Introduction

The self-organizing and cooperative operation of mobile and wireless nodes within ad hoc networks bears several interesting research challenges, of which routing is very prominent. In this area, the main directions of research include performance optimizations and scalability issues. Recently, quality of service and security have also drawn attention.

Being designed to operate under a wide variety of circumstances, most protocols silently assume only well-behaving and cooperative nodes to allow for multi-hop operation of the network. When operating outside of laboratory conditions, the possibility of misbehaving nodes arises. The dependability of the routing system, namely reliability, resilience and fault tolerance under these unfriendly conditions needs to be addressed. Currently, no analytical model exists that describes the effect of misbehaving nodes on the performance of the entire network.

Our investigation provides:

- The detailed classification of node misbehavior as well as a generalized classification to suit analytical models.
- An analytical model of various classes of node misbehavior, including inactive, selfish, and malicious nodes.
- The experimental validation of our model.

Our results enable the precise prediction of the effect of node misbehavior on the overall network behavior within ad hoc networks. The models we present as well as the insights we obtain are an important tool which can be applied to develop more dependable routing protocols.[1]

In Section II, we introduce a classification of multiple types of node misbehavior and derive a more specific classification scheme to suit the needs of analytical modeling. Section III details the modeling process of various classes of misbehavior and extends the work in [2]. The model equations presented are additionally validated by means of simulation. Section IV presents related work. We finish by drawing conclusions and by pointing to possible future work.

II. Classification of Node Misbehavior

There is no common classification of node misbehavior. The authors of [3] and the other related work given in Section IV each introduce their own categories of misbehavior using dissimilar nomenclatures. Since these categories and especially the accuracy of their definition do not suit analytical models like [2], we need to classify the misbehavior differently. An intuitive model of node misbehavior incorporates a lot of different alternative actions a node may perform. From a technical perspective, these degrees of freedom may be implemented as follows:[2]

- *Time*, the on/off behavior of a node may be characterized using {start time, stop time}.
- *Degree of behavior*, giving the probability with which the node behaves as specified {p}.
- *Plane of behavior*, controlling which part of the protocol is affected {control plane, data plane, both}.
- *Type of behavior,* determining which action to perform {forward packet, discard packet, inject packet}.
- *Behavior against whom*, which nodes are affected from the behavior {all nodes, a subset of nodes, a superset of nodes, none}.

Moreover, the misbehavior may occur at different layers. For our investigation, we implemented the above mentioned flavors of behavior within the Qualnet® network simulator. For reasons of complexity we omitted “selective” malicious nodes, which only act maliciously against subsets of all nodes. Given the sheer complexity of the intuitive approach towards node misbehavior, we additionally characterized node misbehavior using some well-defined classes to allow for further analytical study. Our class-based approach aggregates the types of node behavior, which should, on the one hand, be analytically tractable, while, on the other hand, model realistic behavior. Here is a non-exhaustive list of the derived classes:

- C*ooperative nodes*, which comply to the standard, at all times.
- *Inactive nodes*, which include *lazy nodes* (unintentionally misconfigured) and *constrained nodes* (e.g. energy-constraint or field-strength-constraint).

[1] We assume basic knowledge of the concepts underlying the AODV protocol [1]. An analytical model of the AODV route acquisition process is described in detail in [2] and serves as basis for this work.

[2] Please note, that these behavior sets are not necessarily orthogonal to each other and that arbitrary combinations may not make much sense.

- *Selfish nodes*, which optimize their own gain, with neglect for the welfare of other nodes.
- *Malicious nodes*, which inject false information and/or remove packets from the network.

We note that, depending on the degree of non-cooperation the nodes exhibit, selfishness may partially overlap with inactivity. Further restrictions to the classes are outlined in the corresponding sections below.

III. Modeling of Node Misbehavior

Our model of node misbehavior is based on an idealized model of the route acquisition process executed by the AODV protocol which is presented in [2]. This model allows to predict the probability density function of estimated route lengths within the network. This metric describes the statistical relation between the distance of two nodes inside the modeled area and the corresponding probability of being connected. See [2] for the exact derivation of the base model. Important variables include the distance d between source and destination, which may be expressed using the hopcount h of the shortest path between these nodes. In combination with the node degree of the network M and the distribution of node positions we can derive the probability density $p(d)$ and the corresponding probability distribution function $P(d)$ which give the route length distribution inside the network.

Within this work, we extend the model to cover the effect of node misbehavior as well. We formulate the model for *inactive nodes*, *selfish nodes*, and *malicious nodes.* The deformation of the probability distribution when misbehaving nodes are present allows to characterize the network behavior in comparison with the misbehavior-free case.

A. Inactive Nodes

The behavior of *inactive nodes* can be easily described and traced analytically. In reality, they may be constrained nodes or lazy and misconfigured nodes which are intentionally or unintentionally not actively participating in route discovery and packet forwarding.

Definition: An *inactive node* is neither active on the control plane nor on the data plane. It does not cooperate during the routing process and does not forward any packets.

Our model assumes that *inactive nodes* are neither the source nor the destination of a route. Since our definition of behavior concerns the network layer, these nodes may operate on layer 2. We assume that *inactive nodes* do not cause errors on the layers below the network layer. Within our model, *inactive nodes* are extracted from the network. The number of nodes is effectively decreased by the number of *inactive nodes*. Let the proportion of *inactive nodes* be q_{in} and the total number of nodes be n. The number of *inactive nodes* is then $q_{in}n$ and the number of active nodes $(1-q_{in})n$. Only the active nodes participate in the route discovery cycle.

As expected, the node density decreases as the number of *inactive nodes* increases. The average number of nodes within a transmission radius is given by the node degree M. Using the results of [2] we obtain:

$$r'_0 = \sqrt{\frac{A}{\pi n(1-q_{in})}} = \frac{1}{\sqrt{(1-q_{in})}}\sqrt{\frac{A}{\pi n}} = \frac{r_0}{\sqrt{(1-q_{in})}} .$$

The node degree, M', is thus $M' = (r/r'_0)^2 = (r/r_0)^2(1-q_{in}) = M(1-q_{in})$ if we consider inactive nodes, the so called normalized radius being $r'_N = \sqrt{M'}$. As long as the network is sufficiently connected, normal operation will be possible. With increasing number of inactive nodes and decreasing density respectively, the network gets partitioned and the communication will be restricted to subsets of nodes. See Figure 1 and Figure 2 for a visualization of the tests with inactive nodes. The experimental validation within a 500 nodes setup with initially $M = 9$ showed 50 and 100 inactive nodes can easily be tolerated. From 150 *inactive nodes* onwards ($M' = 6.25$) the first drops in connectivity are visible. For 200 ($M' = 5.38$) the routing is significantly burdened, while increasing the number to as high as 250 renders the network nearly disconnected.

B. Selfish Nodes

Selfish nodes maximize their own gain. They do not aid other nodes on the data-plane, thus actively discarding packets routed through them. On the other hand, to be able to send and receive packets from other nodes, they are cooperative on the control-plane, namely the routing process.

Definition: A *selfish node* does not forward any data packets for other nodes except for himself. He cooperates during the route discovery cycle to maintain a concise routing table and to be present in other routing tables.

Due to *selfish nodes*, routes that exist (cooperation for route discovery) may not be used to relay any packets to the destination (non cooperation for data packets). Within standard AODV neither source nor destination are able to detect this misbehavior. From a destination perspective there is an active

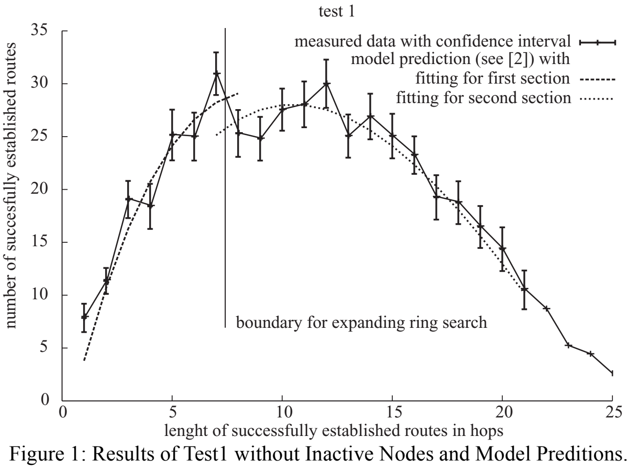

Figure 1: Results of Test1 without Inactive Nodes and Model Preditions.

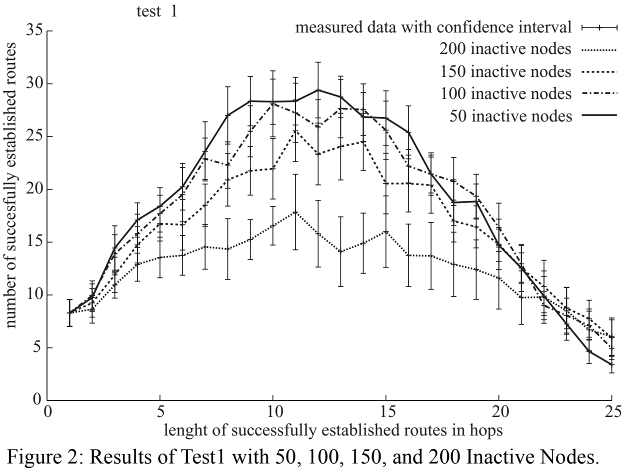

Figure 2: Results of Test1 with 50, 100, 150, and 200 Inactive Nodes.

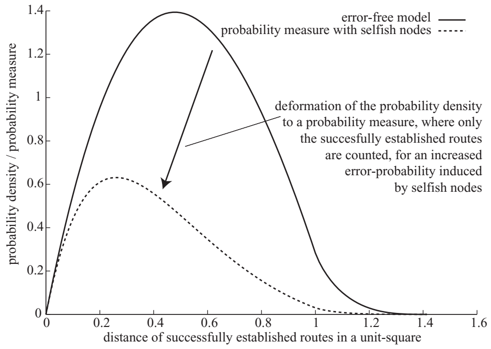

Figure 3: Example for Probability Measure Function with Selfish Nodes.

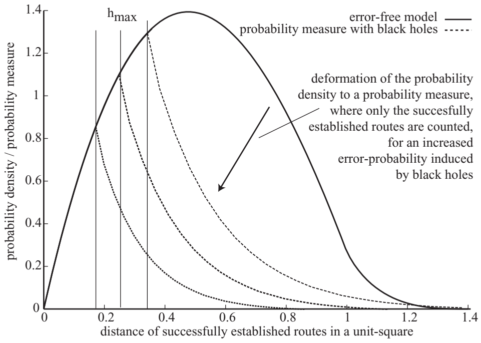

Figure 4: Example of Probability Measure Function with Black Holes.

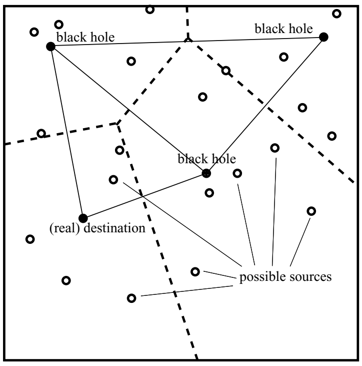

Figure 5: Sphere of Influence of three Black Holes and one Real Destination.

route. However, it is not possible to predict which packets arrive using this route. From the source perspective, the application level packets are sent, but there is no reply. Tracing the data packets inside the network is also then not possible.

We modify our model equations to describe the selfish behavior as follows. If we insert a fraction q_{dp} of nodes which do not forward data packets, we obtain an error probability of q_{dp} for neighboring nodes using a point-to-point connection. The errors on layer two and below may also add some additional loss. In the absence of *selfish nodes* $1 - q$ is the success probability. Since the errors induced by *selfish nodes* are independent of link layer errors, the combined success probability is $(1 - q)(1 - q_{dp})$. This holds for neighboring nodes except when the packets are sent to the *selfish node* itself. If we further exclude collusions among nodes, this probability holds for each node independent of predecessors. For h hops, the resulting probability of a successful data packet transmission is $((1 - q)(1 - q_{dp}))^h$. To be precise, we would need to correct this term using the number of data streams to and from the *selfish node*. We neglect this additional factor without loss of generality.

Combining of the resulting probability with the results from [2], we obtain a modified function for the probability measure, which now gives the estimated number and probability of routes which carry data streams without errors. This does not include the number of selfish routes or the estimated number of discarded data packets. See Figure 3 for the estimated deformation of the curve for a probability of 10% *selfish nodes*. Since our *selfish nodes* take part in the route discovery cycle, the distribution of routes does not change, reflecting the non transparency of *selfish nodes* to all other nodes. The behavior of this sort of nodes is more severe to the network than the behavior of *inactive nodes*.

C. Malicious Nodes

Malicious nodes reduce the utility of the network, without regard for their own gain. Maliciousness may naturally take on many forms. We choose the notion of *black holes*, which masquerade with a fake destination and thus try to attract routes and data packets.

Definition: A *malicious node* abuses the cooperation among nodes to hinder operation of the network.

Definition: A *black hole* answers each route request with a faked route reply claiming to have a one hop route to the destination. If data packets arrive, the *black hole* discards these packets.

Studying the standard AODV behavior, the consequences of *black hole* behavior is obvious. *RREQs* are propagated until a node is or knows of the destination and answers with a *RREP*. The source of the request accepts the first incoming answer and then only the answers with the same or newer destination sequence number and lower hopcount (i.e. shorter and current routes). If the *RREQ* only reaches the intended destination, the *RREP* is correctly accepted by the source and the data transfer should also be successful. If the *RREQ* only reaches one or multiple *black holes*, the source sends data towards one of these. Normal protocol operation assumes only one destination node. Introducing *black holes* changes this behavior. The *black hole* acts as data sink, announcing itself as being one hop away from a fake destination. This may be described as multiple concurrent destinations. Due to the protocol operation of AODV, the node with the shortest route will win the “competition”. We can model this behavior using the areas dominated by *black holes* vs. the area dominated by the original destination. A source will only obtain a valid reply if it is located in the sphere of influence of the valid destination. The distance may thus serve as a metric to describe the influence of *black holes*.

Let us assume only one *black hole* within the network, that all nodes are randomly placed, and that the number of nodes is very large. In this simple case, we can use the sphere of influence to illustrate the behavior. Figure 5 shows a 25 node example. The *black hole* effectively separates the network into two areas. All nodes closer to the *black hole* than to the destination will be trapped. The border between the areas is given by the perpendicular bisector between the *black hole* and the real destination. The catchment area is restricted to the $(n + 1)$ th part of the simulation area, n being the number of *black holes*. See Appendix A for the mathematical proof of this relation.

Since our initial assumptions include the random placement of nodes, we determine the number of nodes inside the individual catchment areas to be approximately $1/(n+1)$ of all nodes. The consequences for the route discovery process are devastating. A successful transmission of data is only possible if the source node is located in the catchment area of the intended destination node. Calculation of the estimated distance from a *black hole* gives (see Appendix B for the calculation, the symbolic figures used to represent the shape of the sphere of influence are a circle and a square):

$$h_{max} = \frac{d_{max}}{r_1} = \frac{1}{r_1}\sqrt{2A}\,,\; h^{\Box}_{max} = \frac{1}{r_1}\sqrt{\frac{2A}{n+1}} = h_{max}\sqrt{\frac{1}{n+1}}\,,$$

$$h^{\circ}_{max} = \frac{1}{r_1}\sqrt{\frac{4A}{\pi(n+1)}} = h_{max}\sqrt{\frac{2}{\pi(n+1)}}\,.$$

As a consequence, all destinations farther than $h_{max}/2$ away from the source will probably be *black holes*. The deformation of the resulting probability measure curve will be very strong. For distances greater $h_{max}/2$, the number of valid routes will heavily decrease since a *black hole* will most probably answer the *RREQ*. Figure 4 depicts a qualitative estimate of the probability measure function. As a result, we see that even a few *black holes* may hinder large areas of the network being connected. *Black holes* are able to inflict far more damage than the other types of misbehavior we discuss.

1) Experimental Validation for Malicious Nodes: The experiments to validate the *malicious node* model are Test2 for pure AODV and Test3 for Gossip enhanced AODV (see Table 1 in Appendix C for the simulation parameters). In order to quantify the impact on the data-plane, we simulated 25 continuous CBR streams. Using a rate of 4 packets/sec of size 512Byte, we obtain a rate of 2kByte/s. We used a stationary scenario for our simulation.

Figure 6 and Figure 7 show the impact of as few as 2% of *black holes* (10 from 500 nodes) for AODV and Gossip enhanced AODV respectively. The predicted decrease is noteworthy and can be seen in both figures. Comparing the numerical result with our prognosis we obtain:

$$h^{AODV}_{max} = \frac{\theta_{AODV}}{r_1}\sqrt{2A} = \frac{1.064}{176.679}\sqrt{2(3742.92m)^2} \approx 31.8\,,$$

$$h^{\Box}_{max} = h_{max}\sqrt{\frac{1}{n+1}} = 31.8\sqrt{\frac{1}{11}} \approx 9.6\,, \text{ and}$$

$$h^{\circ}_{max} = h_{max}\sqrt{\frac{2}{\pi(n+1)}} = 31.8\sqrt{\frac{2}{11\pi}} \approx 7.67\,.$$

Please note that the initial parameterization of our model is described in detail in [2]. The results confirm the prediction that the drop would occur around $h_{max}/2$ which gives $h = 4$ or $h = 5$ as also seen in the simulation. The calculation of the loss the *black hole* introduces is performed using: $d = h(d)r_1$ and $h^{\circ}_{max}/2 \approx 3.8 \Rightarrow d = 0.181$. Integration of the probability measure function gives:

$$P(d) = \int_0^{0.181} p(d)dd = \int_0^{0.181} (1-q_{ers})^{\frac{d\theta_{ers}}{r_1}} 2d(d^2-4d+\pi)dd$$
$$\approx 0.08217 = 8.22\%$$

This proved an accurate prediction for our experimental results which are: $P(d) = 0.0811$ with standard deviation $\sigma = 0.04462$. Further experimental results to illustrate the influence of different types of misbehavior in amore macroscopic fashion cannot be presented due to space constraints.

IV. Related Work

Most ad hoc routing protocols consider the existence of non protocol conformant nodes to be only of minor importance. Firstly, there is some general work in the area of ad hoc networks and security, of which Zhou and Haas [4] and Hubaux et al. [5] are prominent. These elucidate common problems and threats related to ad hoc networks. Solutions for some of the discovered problems can be found in various works including [6], [7], [8], and [9]. Herein, a common approach towards secure ad hoc routing protocols is the use of cryptographic mechanisms to secure the ad hoc routing process. In addition, there exist some related work that tries to mitigate the misbehavior of nodes including [10] and [11]. These existing approaches to secure ad hoc networks build on differing prerequisites, ranging from a single security association between the corresponding nodes to the assumption of an always available public key infrastructure to support operation. The effects of node misbehavior, against which some of the named proposals are targeted, have not yet been well described. Some work, such as Michiardi and Molva [3], [12] describe the influence of misbehaving nodes. The underlying simulation approach, however, cannot be easily generalized. One thing that is missing in literature is an analytical description of the effects that misbehaving nodes induce.

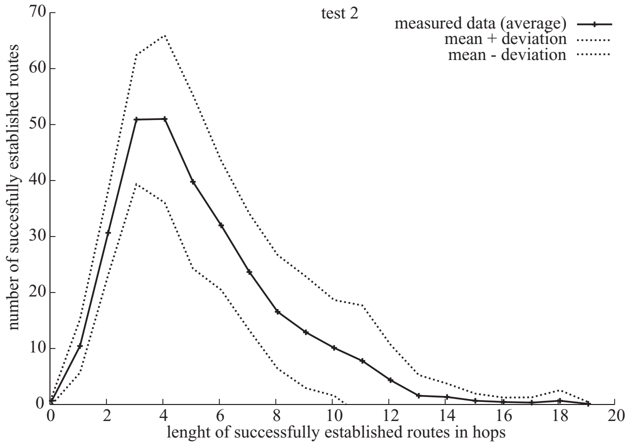

Figure 6: Results of Test2 (AODV) with 2% Black Holes.

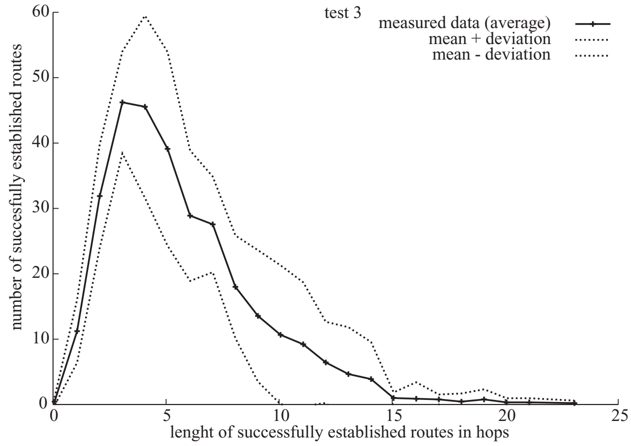

Figure 7: Results of Test3 (Gossip enhanced AODV) with 2% Black Holes.

V. Conclusions

We have discussed the effects of node misbehavior in ad hoc networks. Starting with a general and intuitive classification of node misbehavior, we derived well-defined classes of misbehavior suitable for analytical study. An analytical model covering the different types of misbehavior was presented and adjoined to an existing analytical model of the idealized route acquisition process within AODV [2]. To gather insights on the effects of misbehaving nodes, the estimated impact of these nodes on the overall routing performance was traced analytically as well as validated by means of simulation. As a result we show that *inactive nodes* only moderately harm ad hoc networks, while *selfish nodes* and *black holes* may have devastating influence on the routing process.

The promise of ad hoc networks is built upon the premise of cooperation among nodes. We have shown the network frailty in the absence of such a cooperation. The insights and the models presented assist protocol designers in developing more dependable network protocols. This includes the use of realistic assumptions about potential node misbehavior. As future work, we perceive the improvement of currently available routing protocols with respect to the reliability and availability of their operation.

VI. References

[1] C. E. Perkins, E. M. Belding-Royer, and S. R. Das. Ad hoc on-demand distance vector (aodv) routing. Experimental, RFC 3561, July 2003.

[2] M. Hollick, J. Schmitt, C. Seipl, and R. Steinmetz. The ad hoc on-demand distance vector protocol: an analytical model of the route acquisition process. In Proceedings of Second International Conference on Wired/Wireless Internet Communications, WWIC 2004, Frankfurt (Oder), Germany, pages 201–212. February 2004.

[3] P. Michiardi and R. Molva. A game theoretical approach to evaluate cooperation enforcement mechanisms in mobile ad hoc networks. In *Proceedings of the Conference on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt'03)*, March 2003.

[4] L. Zhou and Z. J. Haas. Securing ad hoc networks. *IEEE Network Magazine*, 13(6):24–30, 1999.

[5] J.-P. Hubaux, L. Butty·n, and S. Capkun. The quest for security in mobile ad hoc networks. In *Proceedings of the 2001 ACM International Symposium on Mobile ad hoc networking & computing (MobiHoc2001)*, pages 146–155. ACM Press, October 2001.

[6] P. N. Seung Yi and R. Kravets. Security-aware ad-hoc routing for wireless networks. Technical Report UIUCDCS-R-2001-2241, University of Illinois at Urbana-Champaign, Department of Computer Science, August 2001.

[7] Y.-C. Hu, D. B. Johnson, and A. Perrig. Sead: Secure efficient distance vector routing in mobile wireless ad hoc networks. In *Proceedings of Fourth IEEE Workshop on Mobile Computing Systems and Applications (WMCSA '02)*, pages 3–13, June 2002.

[8] M. G. Zapata and N. Asokan. Securing ad hoc routing protocols. In *Proceedings of the ACM workshop on Wireless security (WiSe2002)*, pages 1–10, August 2002.

[9] P. Papadimitratos and Z. Haas. Secure routing for mobile ad hoc networks. In *Proceedings of the Conference on Communication Networks and Distributed Systems Modeling and Simulation (SCS 2002)*, pages 27–31, January 2002.

[10] S. Marti, T. J. Giuli, K. Lai, and M. Baker. Mitigating routing misbehavior in mobile ad hoc networks. In *Proceedings of the Sixth Annual IEEE/ACM International Conference on Mobile Computing and Networking (MobiCom2000)*, pages 255–265, August 2000.

[11] S. Buchegger and J.-Y. L. Boudec. Nodes bearing grudges: Towards routing security, fairness, and robustness in mobile ad hoc networks. In *Proceedings of the Tenth Euromicro Workshop on Parallel, Distributed and Network-based Processing*, pages 403–410, January 2002.

[12] P. Michiardi and R. Molva. Simulation-based analysis of security exposures in mobile ad hoc networks. In *Proceedings of the Conference on Next Generation Wireless Networks: Technologies, Protocols, Services and Applications (European Wireless 2002)*, February 2002.

Appendix A. Sphere of Influence of a Black Hole

Assume a network which consists of exactly one destination node and exactly one *black hole* which serves as a fake destination node, the positions of these nodes being i.i.d. Each possible configuration is complemented by exactly one opposite configuration of *black hole* and destination node. Likewise, the sphere of influence from the real destination node and the fake destination node are interchangeable. We denote these areas A_1 and A_2. We calculate the area which is dominated from one particular node as $\bar{A}$: $\bar{A} = (A_1 + A_2)/2 = A/2$ with $A_1 + A_2 = A$.

The generalization to cover one destination node and n *malicious nodes* gives $(n+1)!$ constellations. Each node can appear at $(n+1)$ positions, which it occupies $n!$ times. The mean area $\bar{A}$ is given by:

$$\bar{A} = \sum_{j=1}^{n} n! A_j \cdot \frac{1}{(n+1)!} = \sum_{j=1}^{n} A_j \cdot \frac{1}{n+1} = \frac{A}{n+1}.$$

Especially the concentration of *black holes* in certain areas may lead to other results. The areas are moreover not necessarily of the same shape. Considering these boundary conditions, the formula will generally hold.

Appendix B. Influence of Black Holes on h_{max}

As shown above, the generalized sphere of influence of a *black hole* is $A/(n+1)$. An area-equivalent square will have a side length of $\sqrt{A/(n+1)}$. We obtain an estimate for the distance h_{max}. The maximal length inside the square is the diagonal line with length $\sqrt{(2A)/(n+1)}$. The maximum hopcount is then:

$$h^{\square}_{max} = \frac{1}{r_1}\sqrt{\frac{2A}{n+1}} = h_{max}\sqrt{\frac{1}{n+1}}.$$

Under the assumption that the covered area is represented better by a circular area than a rectangle, we can transform the result into an area-equivalent circle. The radius being $\sqrt{A/(\pi(n+1))}$. The maximum distance equals the diameter and thus the hopcount is:

$$h^{\circ}_{max} = \frac{1}{r_1}\sqrt{\frac{4A}{\pi(n+1)}} = h_{max}\sqrt{\frac{2}{\pi(n+1)}}.$$

These results now can easily be compared to the error free case of h_{max}: $h_{max} = d_{max}/r_1 = (1/r_1)\sqrt{2A}$.

Appendix C. Experimental Parameter Set

Table 1 gives the experimental parameter set used for the simulations.

Table 1. Experimental Parameter Set

Test / Variable	**Test1 AODV**	**Test2 AODV**	**Test3 AODV + Gossip**
Simulation Area	$(3300.94\text{m})^2$	$(3742.92\text{m})^2$	$(3742.92\text{m})^2$
Replications	20 each	10	10
Mobility	no	no	no
Gossip (p,k)	-	-	(0.7, 1)
ERS	no	yes	yes
Traffic	each 10s one stream	each 100ms one stream	each 100ms one stream
Packets (in Flow)	1	4 packets /s	4 packets /s
r_0	83.287m	94.438m	94.438m
r_1	176.679m	176.679m	176.679m
M	9, M' varies	7	7
Node Behavior	inactive nodes	2% black holes	2% black holes
Common Parameter Set for all Simulation	Number of Nodes = 500; Duration = 500s, Transmission Power = 7dBm; Propagation Model = Free Space Transmission Range (r) = 249.862m; r_1= 176.679m; MAC 802.11b; Max. Transmission Rate = 11 MBits/s Local Repair = Deactivated; Hello Messages = Deactivated; Packet Size = 512Byte; UDP as Transport		