Exploiting Distributed Source Coding for Multi-hop Routing in Wireless Ad Hoc Networks

Sebastian Kühlmorgen, Andreas Festag, Gerhard Fettweis
Vodafone Chair Mobile Communications Systems, Technische Universität Dresden, Germany
{sebastian.kuehlmorgen, andreas.festag, gerhard.fettweis}@tu-dresden.de

Abstract—This paper presents a routing algorithm for wireless multi-hop ad hoc networks, which applies distributed source coding (DSC) and multi-path transport of data packets. DSC enables the relay nodes to efficiently encode the packets and helps the destination node to correctly decode the information from multiple erroneous copies of the same packet received via different paths. While in conventional communication systems packets with bit errors are discarded, in our design the relays are allowed to forward erroneous packets. In order to exploit DSC for ad hoc routing, we extend contention-based geographical forwarding (CBGF) to support multi-path packet transport and enable the joint decoder to recover transmission errors. Under harsh wireless conditions, the approach promises performance gains compared to the case without joint decoding. In order to evaluate the performance of CBGF, we define an evaluation framework that models the erroneous forwarding and joint decoding by an abstraction of the physical transmission. We assess the performance by simulations for a simple scenario with static nodes in terms of packet success ratio, end-to-end delay, and data traffic overhead, and compare the results with the baseline CBGF scheme without DSC-specific extensions and joint decoding.

I. INTRODUCTION

Distributed source coding (DSC) is a technique where nodes encode correlated data independently from each other. A single destination node decodes the combination of data received from several other nodes via different links; this process is referred to as joint decoding. Considering joint decoding the Slepian and Wolf theorem [1] has theoretically shown that separate coding is as efficient as joint coding for lossless compression.

The concept of DSC can be applied to wireless sensor networks, where several sensors send their compressed measurement data to a central point at a lower energy consumption compared to a solution without joint decoding [2]. The application of DSC can be generalized to ad hoc networks with short range wireless technologies, where data are transferred via multiple wireless hops towards the destination. Typically, wireless multi-hop ad hoc networks are designed such that a data packet is forwarded via a single predefined network path from a source to a destination via relay(s). Our proposed scheme specifically addresses scenarios, where multiple simultaneous network paths are employed in order to increase the reliability and robustness of packet transport. Unlike in conventional communication systems, we explicitly allow the forwarding of erroneous packets instead of discarding them.

The joint decoding of packets at the destination facilitates to recover the original information from erroneous packets. The bit errors may occur at any of the previous wireless links of the network path. At the destination, the combination of the corrupted packets may help the correct decoding, increase the likeliness of reception and render end-to-end retransmissions unnecessary. Overall, the joint decoding of packets received via different paths promises to improve the robustness and reliability of transmissions at the expense of redundantly transmitted packets increasing the load in the network.

We study the application of DSC in the context of contention-based geographical forwarding (CBGF). CBGF utilizes geographical positions for efficient forwarding at low control overhead, which avoids expensive route establishment and maintenance procedures of topology-based routing schemes. In order to exploit DSC in the forwarding process we introduce functional and performance-related enhancements into the original algorithm, including multi-path routing and a forwarding logic that uses a confidence indicator as measure of distortion of the packet in addition to the geographical progress of the packet. We evaluate the performance of the DSC-enhanced forwarding algorithm by simulation in a simple scenario with static nodes and use network-specific metrics, i.e. packet success ratio, end-to-end delay and data traffic overhead. The evaluation requires an accurate yet efficient model of the physical transmission. We present an evaluation framework based on an abstraction of the physical transmission, called blackbox, which models the joint decoder.

The contributions of the present paper are as follows: (i) We present protocol design enhancements for CBGF to exploit DSC. (ii) We describe a framework for simulation-based performance evaluation that is based on a physical layer abstraction and allows for accurate yet efficient performance assessment. (iii) We provide results of a simulation-based performance evaluation that quantifies the performance gain of the DSC-enhanced forwarding scheme compared to the conventional variant without DSC. Previous publications address DSC from an information theoretic perspective, consider generic decode-and-forward (DF) algorithms, or apply DSC in specific application scenarios, such as sensor networks. Compared to these approaches we apply the DSC approach to ad hoc networks and study the applicability of DSC for a state-of-the-art routing protocol.

The remainder of the paper is organized as follows: Section 2 provides background on DSC, followed by a description of protocol enhancements to enable DSC for CBGF in Sec. 3. Sections 4 and 5 present the evaluation framework and the performance results, respectively. Section 6 concludes the paper.
a source node, which sends a data packet to $D$ overheard by the relay node $u$.

**Fig. 2. Joint decoding of Slepian-Wolf theorem.**

Admissible rate region  

$H(u_r | u_s)$  

$H(u_s | u_r)$  

$H(u_s)$  

$H(u_r)$  

Region A is the rate when the packets are uncorrelated and region B is the extension enabled by the Slepian-Wolf theorem.

**II. INFORMATION THEORETIC BACKGROUND**

We briefly present the Slepian-Wolf correlated source coding theorem [1] as a basis for DSC and its application in the forwarding algorithm.

Let $S$ and $R$ be a pair of correlated sources, which separately encode data and send the data to a destination $D$ for joint decoding. The theorem defines the bounds for the information rates $R_s$ and $R_r$, which ensures lossless decoding at an infinite code length:

$$R_s \geq H(u_s | u_r),$$  \hspace{1cm} (1a)

$$R_r \geq H(u_r | u_s),$$  \hspace{1cm} (1b)

$$R_s + R_r \geq H(u_s, u_r),$$  \hspace{1cm} (1c)

where $H(\cdot)$ represents the entropy function, $R_s, R_r$ are the rates between $S - D$ and $R - D$, respectively, and $u_s, u_r$ are the data generated by $S$ and $R$. The achievable rate region is depicted in Fig. 1. The achievability of the bounds has been generalized to arbitrary ergodic processes, countably infinite alphabets, and an arbitrary number of correlated sources [3].

Transforming the correlated source coding problem into a simple network of three nodes (Fig. 2), $S$ can be regarded as a source node, which sends a data packet to $D$. This packet is overheard by the relay node $R$, which re-encodes and forwards the data. The destination node $D$ performs joint decoding of the two received data packets. The Slepian-Wolf theorem determines the minimum information rate at which $S$ and $R$ need to encode the packet such that $D$ can correctly, i.e. error-free, decode it.

While information-theoretic studies target at finding the minimum rate and maximum compression, practical aspects are related to code design, selection of code rates, and low-complexity decoding. However, instead to determine theoretical limits the DSC approach, in this paper we are rather interested in the performance gain of a state-of-the-art and DSC-enabled DF algorithm.

**III. DSC-ENHANCED FORWARDING ALGORITHM**

Contention-based geographic forwarding (CBGF) is a routing algorithm for ad hoc networks [4] that utilizes geographic positions for wireless multi-hop forwarding. Compared to topology-based routing, where routes are established on demand or reactive, in CBGF the nodes forward packets using their local knowledge; hence CBGF does not require the establishment and maintenance of routes. CBGF has been studied in wireless sensor networks, vehicular communication networks, and others.

CBGF is a receiver based forwarding algorithm where the receiver decides if the packet will be forwarded or not based on overhearing. The advantage is – compared to sender based forwarding – that no beacons are necessary to get the neighbor position information since the node broadcasts the packet to all neighbors. After reception, the node buffers the packet and starts a timer depending on the distance among the sender, relay and destination. The node closest towards the destination has the shortest timer. After the timer expires, the packet is re-broadcasted which is overheard by the neighbors. When the packet is received a second time in the contention period the node stops the timer and drops the packet. Also packets with negative progress are discarded.

In order to enable DSC in an ad hoc network, several changes and extensions need to be applied to CBGF, which are explained next:

**Multi-path distribution.** DSC applies joint decoding of different copies of the same packet received at the destination in order to recover the originated packet from the source. Therefore, enhanced CBGF has to provide multi-path dissemination of packets. This is achieved by enabling relay nodes to re-broadcast packets and by introducing a retransmission counter (RC). After the reception of an additional copy in the contention period RC is incremented up to a pre-defined retransmission threshold (RT); when RC exceeds RT the packet is dropped. RC and RT allow to limit the number of redundant re-transmissions and indirectly the number of multiple paths, and control of the network load in densely populated networks.

**Selective forwarding of packet copies.** Joint decoding requires uncorrelated errors in the packets and therefore the different network paths have to be disjoint. For this reason the addresses of the last two senders are carried in the packet header and provided to the next hop (how this is achieved see Sec. IV). If another node forwards a packet, which the current forwarder has already received from the same sender, the node will recognize it and forward a different copy. Moreover, the packet from the source is assumed to be the best copy and will be forwarded every time.

**Enhanced CBGF timer determination.** The timeout of the CBGF timer is calculated based on the spatial progress of the packet, which cause some issues: In dense scenarios, potential relays are likely located at the edge of the communication
range and in close proximity, which leads to almost equal timeouts and in two ways inappropriate choices of forwarded packet copies. First, in our scheme we allow lossy forwarding and if the farthest relay forwards the packet, this packet will likely have the most errors. Second, if the timers at close relays expire at almost the same time the packets are passed to the MAC queue and no overhearing will take place anymore with the result that false, i.e. copies with correlated errors, will be forwarded. To avoid these issues, a confidence value that describes the distortion of the packet, estimated at the PHY and provided to the NET, is taken into account in the calculation of the timer. Overall, the timeout calculation is split into two parts: progress \( (pr) \) and confidence value \( (cv) \) and is determined by the following equations:

\[
t_{pr} = t_{max,pr} - \frac{t_{max,pr} - t_{min,pr}}{d_{max}} pr, \quad (2a)
\]

\[
t_{cv} = t_{max,cv} - t_{min,cv} (cv_{max} - cv_{min}) + t_{min,cv}, \quad (2b)
\]

\[
t_{to} = t_{pr} + t_{cv}, \quad (2c)
\]

\( t_{to}, t_{pr} \) and \( t_{cv} \) are the overall timeout, confidence value timeout and progress timeout, respectively. Max and min represent the maximum and minimum values and \( d_{max} \) is the maximum expected communication range.

Avoidance of incorrect discarding. With regular CBGF, a node decides to discard a packet when it determines that the forwarding progress is negative. Such situation may occur when a node forwards a packet and the neighbors of the receiving node did not. When now the receiving node forwards its packet the neighbor nodes with negative progress drop it. In addition, when these neighbor nodes receive another copy of the packet, they recognize it as duplicate and discard it as well. Instead of discarding the packet, these neighbors may also contribute to reliable packet delivery. While in regular CBGF it belongs to the normal operation, in enhanced CBGF it aggravates due to the multi-path extensions. Therefore, we modify the duplicate packet detection mechanism, such that packets with negative progress are removed from the duplicate detection list, allowing subsequent packets to be forwarded.

IV. EVALUATION FRAMEWORK

We present the evaluation framework used for the evaluation, including the scenario, the protocol stack, and the abstraction of the physical transmission.

A. Scenario

For performance assessment of enhanced CBGF we define a simple scenario with four or five nodes, respectively, i.e. one source \( S \), one destination \( D \), and two or three relays \( R_i \) between \( S \) and \( D \) (Fig. 3). The relays are located closer to the source than to the destination and placed at 30% of the distance \( d \) between \( S - D \). This setup ensures that \( R_i \) will normally receive the packet from \( S \) and the \( SINR \) between \( R_i \) and \( D \) is low. The links between \( S \) and \( R_i \) are assumed to be lossy (indicated by the bit error probability \( p_i \) of link \( i \) in Fig. 3) and the relays forward packets also when they are corrupted. Depending on the value of \( d \), two cases can be distinguished: (i) either \( S \) reaches \( D \) directly and the packets forwarded by \( R_i \) support the decoding at \( D \), or (ii) no direct link between \( S \) and \( D \) exists and \( R_i \) forward the packet to \( D \) for joint decoding.

The scenario extends the network model of the Slepian-Wolf theorem (Sec. II) since it introduces a single source and relay nodes with links between \( S \) and \( R_i \).

B. Protocol Stack

The nodes execute a simplified protocol stack composed of layers 1–3 of the ISO OSI model with physical transmission (PHY) and medium access control (MAC) layers, as well as Enhanced CBGF at the network layer (NET) and an application layer (APP) on top that represents data source and sink (Fig. 4). In principle, for the PHY and MAC we assume a WiFi-like system close to the IEEE 802.11a standard (see parameters in Sec. V), but make significant modifications to exploit DSC:

DSC protocol header. We introduce a dedicated PHY protocol header, referred to as DSC header, that carries the relevant header fields (at least, source, last two sender and destination addresses) and is protected by an error detection code (CRC). The DSC header ensures a minimum set of control information that can be assumed to be error-free, considering the fact that errors may also affect the protocol headers and the forwarding of erroneous packets may result in erroneous distribution through the network.

Encoding & Decoding. We apply an accumulator-assisted distributed turbo code (ACC-DTC) [5], a simple iterative decoding technique that utilizes simple convolutional codes
and exploits the error probability in the $S-R$ link as the source-relay correlation, which is estimated at $D$.

**X-layer packet handling.** When the MAC CRC of the received packet fails, $D$ buffers the packet. Upon reception of the next packet copy from another relay, the PHY joint decoder triggers another decoding attempt. For the joint decoding in combination with CRC processing and packet buffering, $D$ realizes efficient cross-layer packet handling and control information exchange.

In order to illustrate the role of the different protocol layers and their interaction, we describe the data flow from source $S$ via relays $R_i$ to the destination $D$ (Fig. 4): **Step 1** When $S$’s APP triggers a transmission, the message is passed to the underlying NET and the MAC, each layer appending its header to the packet and the frame, respectively. The MAC broadcasts a frame to all neighbor nodes, i.e., the relays. We note that the MAC does not acknowledge the transmission of broadcast frames. **Step 2** The relay nodes decode the received packet and execute the NET forwarding procedure, i.e., buffer the packet and starts the CBGF timer. **Step 3** When first timer in a relay expires, the packet, which may contain errors, is passed through the protocol stack – including interleaving and re-encoding at the PHY – and forwarded to $D$. The relay refrains from forwarding if the confidence value of the packet (see Sec. III) is below a pre-defined threshold $c_{\text{min}} = 0.7$, which corresponds to $20\%$ of errors inside the packet. **Step 4** At the destination $D$, the packet’s MAC CRC is verified; if it fails, the MAC provides this information to the joint decoder. **Step 5** Once another packet copy arrives, the joint decoder re-attempts the decoding of the packet. As soon as the decoding succeeds, the correct packet is passed to the APP.

### C. Physical Layer Abstraction

In order to reduce the system complexity, the evaluation framework uses a PHY abstraction, referred to as blackbox [6], [7], which models the behavior and performance of the joint decoder. In general, an abstraction allows to map the link quality to a single scalar metric, i.e., it is essentially a function that maps a sequence of the varying $\text{SINR}$ values to a corresponding effective $\text{SINR}$ [8] (referred to as Exponential Effective SNR Mapping, EESM) or as an additional approach to a mutual information (MI) [9] [10]. In our case the $\text{SINR}$ and the $p_i$ values from the relays are mapped to the mutual information. In a second step, the effective $\text{SINR}$ or MI is used to find an estimate of the block error probability. Essentially, the blackbox output enables the forwarding algorithm to determine whether a packet is correctly received by using a two stage procedure, replacing complex and computationally intensive calculations.

In detail, the blackbox consists of a deterministic and a stochastic block, i.e. blackbox 1 and 2, respectively (Fig. 5). **Blackbox 1** calculates the mutual information of the packets processed by the joint decoder and passes the estimated MI value to Blackbox 2. Input parameters are code rate, modulation order, bit error rates $p_i$ at the relays for the $S-R_i$ links, and the $\text{SINR}$ for the $R_i-D$ links (Fig. 3). In other words, blackbox 1 maps the $p_i$ and $\text{SINRs}$ to a single value MI, effectively describing the quality of the received block after decoding [6]. **Blackbox 2** provides the number of errors in a packet using the input parameters MI, the number of received packet copies, and the code block length following the distribution:

$$\alpha \delta(0) + (1-\alpha) \mathcal{N}(\mu, \sigma^2),$$

(3)

where $\alpha$ and $(1-\alpha)$ are the fraction of error-free and erroneous packets, respectively, $\mathcal{N}(\mu, \sigma^2)$ is a Gaussian distribution with mean $\mu$ and variance $\sigma^2$. Blackbox 2 is implemented as a look-up table, for which the values were obtained by simulations with the implemented joint decoder for given $\text{SINRs}$ and the results were fitted according to (3). The look-up table only realizes discrete parameters and therefore, the resulting errors need to be determined by interpolation of MI using $\alpha$, $\mu$, and $\sigma$. Moreover, we have restricted the look-up table to three packet copies and a code block length of $1,000$ bytes, respectively. If more than three packet copies arrive at the destination, only the three packets with the best $p_i$ values are taken.

### V. Performance Evaluation

This section presents performance metrics and performance results. For comparison of the baseline and enhanced CBGF, we define the following metrics that assess the network-layer performance in terms of reliability, latency, and network:

#### A. Key Performance Indicators

The packet success ratio $\text{PSR}$ is defined as the number of packets successfully received at the destination divided by the source transmitted packets:

$$\text{PSR} = \frac{\text{No. of Rx packets at destination}}{\text{No. of Tx packets from source}}.$$  

(4)

While the $\text{PSR}$ expresses the reliability of the packet transmission, it does not necessarily indicate the number of packets that are needed to decode the original packet error free. For this reason we introduce the metric ‘division of successfully decoded packets’ $\text{DDSP}$. In case of enhanced CBGF, $\text{DDSP}$ provides the number of packets were needed in the destination’s joint decoder to correctly decode the packet. For comparison, in the conventional decoding we count the number until an error-free packet arrives at the destination. We note that the sum of all individual values corresponds to the packet success ratio $\text{PSR}$.

The average end-to-end delay $\text{E2ED}$ measures the difference of the timestamps when a packet is sent by a source and when it is received at the destination averaged over the delays of all messages:

$$\text{E2ED} = \frac{1}{N} \sum_{i=1}^{N} (t_{\text{Rx},i} - t_{\text{Tx},i}),$$

(5)
whereas $N$ represents the total number of received packets, $t_{Tx,i}$ and $t_{Rx,i}$ are the timestamps when the packet was transmitted and received, respectively.

In order to measure the network load and how often the channel was used, we introduce the data traffic overhead $DTO$. It allows to measure the retransmissions in the network and is defined as the number of transmissions at the PHY layer divided by the number of triggered application messages:

$$DTO = \frac{\text{No. of frames sent at the PHY layer}}{\text{No. of application messages}}$$

B. Simulation Results

For performance evaluation we use the framework presented in the previous section. The simulation is performed with the network simulator 3\(^1\), version 3.19, which was extended by enhanced CBGF, and the DSC-specific extensions of an IEEE 802.11a PHY and MAC protocols. The channel access was performed with the IEEE 802.11a CSMA/CA and its default parameters. The channel is modeled by a log-distance path loss with a channel exponent of 3 and Nakagami fading with shape factor 1. The reception probability is at path loss with a channel exponent of 3 and Nakagami fading was performed with the IEEE 802.11a CSMA/CA and its default parameters. The channel is modeled by a log-distance path loss with a channel exponent of 3 and Nakagami fading with shape factor 1.

The PHY header is encoded with a stronger code and a lower code rate. Therefore, for simplicity reasons and since the header is much smaller than the 1,000 bytes for a packet and the packet is discarded if there are more than 20\% of errors inside we can assume that the PHY header is always error free except a negligible fraction. For this reason the PHY header is not separately modeled in our simulation.

The PSR for base and enhanced CBGF is shown in Fig. 6 (top). We observe that enhanced CBGF outperforms the baseline over the whole distance range for two and three relays. Up to 250 m enhanced CBGF has a PSR of 100\%. For a given distance, the greatest PSR improvement can be found for three relays at 435 m with 42 percentage points. Considering a minimum PSR threshold of 90\% yields a gain of 110 m in distance for 3 relays, and still 72 m for 2 relays. As can be seen in Fig. 6 (bottom), the enhanced CBGF has a shorter average $E2ED$ for all distances than the baseline. The maximum $E2ED$ amounts to 44 ms. Reason for the differences is the fact that in the enhanced CBGF scheme, every packet is forwarded even though there are errors inside or not. In the baseline CBGF, erroneously received packet are discarded and relay nodes rebroadcast these packet when their timer expires, which results in a longer average $E2ED$. Another reason for the shorter $E2ED$ is the fact that the enhanced CBGF algorithm utilizes the link quality to determine the CBGF timer. A good link – indicated by a high confidence value – is associated with a shorter CBGF timer, which in turn reduces the $E2ED$.

It is important to remark that only a few packets arrive at the destination when the distance increases beyond 650 m. Therefore the statistics deteriorate, which becomes visible by the dramatically increasing error bars. For this reason we have increased the number of different seeds beyond 600 m to 200.

\(^1\)Available: http://www.nsnam.org

\[\text{Fig. 6. Packet success ratio PSR (top) and end-to-end delay E2ED (bottom) over distance (d) for two and three relays. Enhanced CBGF (with joint decoding – JD) and baseline CBGF (with normal decoding – NC) is exhibited.}\]

\[\text{Fig. 7. Data traffic overhead DTO over distance (d) normalized by the number of messages sent by the source for two and three relays.}\]

still the results have a high standard deviation. The delays for the two or three relay cases almost overlap since the packet by the third relay is also forwarded in the same temporal range as the first two packet.

Fig. 7 presents the DTO. For the enhanced CBGF the DTO is higher due to the permission of retransmissions of corrupted packets. Furthermore, every packet is forwarded until a predefined retransmission threshold is exceeded (see Sec. III). In the reference CBGF the other nodes refrain from retransmissions if a neighbor already forwarded the same packet. Below 25 m the DTO jumps down to 2 due to the topology of the selected scenario and the fact that both CBGF algorithms do not forward packets with negative spatial progress. While the maximum DTO is 4, we note that our
simple scenario can be considered as a network with a sparse distribution of nodes, which does not lead to a considerable load in the wireless channel. It can be expected that a scenario with a dense node distribution would cause a high channel load. Moreover, in the latter scenario typically a sufficient number of links with higher link quality are available for a correct transmission of the packet, which would cause a highDTO and channel load at a limited or no gain of DSC in PSR and E2ED.

Fig. 8 presents the DSDP over the distance for three relays and the number of packets needed for error-free decoding. In the enhanced CBGF, the packets are merged whereas in the baseline CBGF the destination needs to wait until an error-free packet arrives. For both cases the number of packets is illustrated as follows: square – 1 packet, triangle – 2 packets, circle – 3 packets. The sum of all three graphs are illustrated as follows: square – 1 packet, triangle – 2 packets, circle – 3 packets. We can see that for a distance below 130 m more than 90 percentage points (p.p.) of packets could be correctly decoded with a single packet copy (green and yellow lines). With larger distance, the capability to decode a packet with just one copy decreases and the blue and orange graphs, where two copies are required to correctly decode a packet, grows. At approximately 200 m also the red graph (three packets are needed for decoding) becomes relevant and increases up to 10 p.p. at 420 m. It is worth noting that the yellow, orange and purple line can be treated as baseline, and the difference between the blue and orange graphs and red and purple graphs can be treated as the gain achieved by the joint decoding and enhanced CBGF. Below 330 m the enhanced CBGF graph is below the baseline graph. The reason is that in the enhanced CBGF case the packets were already successfully decoded with two copies whereas in the baseline CBGF the destination has to wait for a third, error-free packet.

VI. Conclusion

In this paper we have presented design aspects for DSC-enhancements of a state-of-the-art geographical routing proto-