Geo-referenced Proximity Detection of Wildlife with WildScope: Design and Characterization

Gian Pietro Picco[1,*], Davide Molteni[1], Amy L. Murphy[2],
Federico Ossi[3], Francesca Cagnacci[3], Michele Corrà[4], Sandro Nicoloso[5]

[1] University of Trento, Italy [2] Bruno Kessler Foundation, Trento, Italy
[3]Edmund Mach Foundation, S. Michele all'Adige, Italy
[4]TRETEC, Trento, Italy [5]DREAM–IT, Pistoia, Italy
[*]Contact author: gianpietro.picco@unitn.it

ABSTRACT

Existing systems for wildlife monitoring focus either on acquiring the location of animals via GPS or detecting their proximity via wireless communication; the integration of the two, remarkably increasing the biological value of the data gathered, is hitherto unexplored. We offer this integration as our first contribution, embodied by our WILDSCOPE system whose key functionality is geo-referenced proximity detection of an animal to others or to landmarks. However, to be truly useful to biologists, the in-field monitoring *system* must be complemented by two key elements, largely neglected by the literature and constituting our other contributions: *i)* a model exposing the tradeoffs between accuracy and lifetime, enabling biologists to determine the *configuration* best suited to their needs, a task complicated by the rich set of on-board devices (GPS, low-power radio, GSM modem) whose activation depends strongly on the biological questions and target species at hand; *ii)* a *validation* in controlled experiments that, by eliciting the relationship between proximity detection, the distance at which it reliably occurs, and the location acquisition, provides the cornerstone for the biologists' analysis of wildlife behavior. We test WILDSCOPE in real-world experimental setups and deployments with different degrees of control, ascertaining the platform accuracy w.r.t. ground truth and comparing against a commercial proximity logger.

Categories and Subject Descriptors

C.2.1 [**Network Architecture and Design**]: Wireless communication

Keywords

Wireless sensor networks, wildlife monitoring

1. INTRODUCTION

Technology and wildlife studies have been increasingly coupled in recent years, to the extent of defining a new discipline called *biologging* [21]. This form of animal-attached remote sensing allows the recording of state variables of the animals (e.g., position, acceleration) and the environment (e.g., light, temperature); animals can therefore be studied in the wild *from their own point of view*. For instance, GPS telemetry provides movement trajectories and position in space through animal-borne GPS devices [4], whose successful application motivates the increasing interest in Movement Ecology [18]. Another technology that only recently revealed its huge potential for wildlife research is radio-based *proximity detection*, where "contacts" among animals are inferred from message exchanges among animal-borne wireless devices, called proximity loggers. These appeared in wildlife ecology less than one decade ago but are rapidly expanding, as they enable scientists to derive rich information on the encounters between individuals, usually obtained by means of social network analysis [14].

The wireless sensor network (WSN) community acknowledged wildlife monitoring as a potential application early on [16, 17]; after a period of quiescence, there is a renewed surge of interest. The concise survey of related work in Section 2, however, shows that the integration of the two powerful state variables of animals above (i.e., position and proximity to other animals) is hitherto unexplored. One reason is that the two corresponding technologies have been developed independently to answer urgent biological questions (e.g., disease transmission [3] in the case of proximity loggers), slanting platform design towards solving immediate needs.

Contribution #1: A geo-referenced proximity detection system. Our first contribution is to reunite these complementary dimensions in a single, integrated system, called WILDSCOPE, providing *geo-referenced proximity detection*, where GPS activations are triggered by contacts. Nowadays, proximity loggers do not provide information about the location where contacts occurred. Instead, GPS loggers provide only location information, typically acquired with a rather large period (e.g., hours) to preserve energy; contacts must be inferred from the intersection of these sparse individual trajectories [9, 22], resulting in poor accuracy. WILDSCOPE overcomes these issues by providing *directly* contact information, as in proximity loggers, and by acquiring location information *when* and *where* a contact occurs. Contacts are no longer *inferred* based on location; on the contrary, locations are acquired when contacts are *detected*. The ability to track the patterns of space usage of an animal *during* the occurrence of a contact enables the investigation of unanswered questions, e.g., the link between the movement patterns of an individual and its encounters.

WILDSCOPE mobile nodes also interact with fixed ones, deployed where the target animals range; in this case, geo-referenced proximity detection enables biologists to monitor, with high spatial and temporal precision, how animals use specific focal resources.

IPSN '15, April 14 - 16, 2015, Seattle, WA, USA.

http://dx.doi.org/10.1145/2737095.2737104.

The omni-directionality of low-power radio and the ability to directly and continuously monitor the animal's visit to a site overcome two major limitations of existing technology, respectively, the limited view field of camera traps and the low time resolution of commonly-used periodic GPS acquisition schedules.

Although conceptually simple, the integration of GPS-based localization and proximity detection is far from trivial, due to many subtle interactions among these functional components. The requirement to enable remote data offloading, in addition to the opportunistic offloading enabled by fixed nodes, further complicates the hardware design of animal-borne nodes, constrained by limitations on shape and weight. The firmware, based on TinyOS, orchestrates the multiple on-board devices (low-power radio, GPS, GSM modem) to realize the integration of GPS and proximity detection, and provide various options for data acquisition and collection. Finally, dedicated tools, working hand-in-hand with the devices, are required in-field, to enable menial and yet fundamental functions such as finding and accessing devices, but also offline, to gather and automatically process the collected data. Section 3 presents the main requirements established by the biologists in our team, co-authors of this paper, while Section 4 illustrates the platform and associated toolset, including design choices that, although coming across as low-level engineering, in practice made the difference between a lab prototype and one able to sustain in-field deployment.

We argue that WILDSCOPE by itself already constitutes an advance of the state of the art, empowering biologists with an observation instrument of unprecedented power. However, for this instrument to be truly useful, biologists must be offered two additional elements that are currently largely neglected by the literature.

Contribution #2: Configuration via a lifetime model. The first element is a *configuration* of the platform best suited to the biologists' research needs. Indeed, their desire to accumulate huge amounts of accurate data must be confronted with the limited energy supply and data storage imposed by collar weight restrictions; an incorrect configuration may interrupt prematurely the acquisition of data from the field. Unfortunately, there is no one-size-fits-all configuration; different species and different biological questions require different configurations, which depend on a combination of technological and biological parameters. The richness of WILDSCOPE complicates matters, as the configuration space is amplified by the many on-board devices whose activation depends strongly on the biological hypothesis and target species being investigated. Therefore, the model we present in Section 5 is key in enabling biologists to estimate the threat to lifetime posed by limited energy supply and data storage. Further, the model is also a useful tool to gain additional insights about WILDSCOPE itself, and understand how it can be further optimized. For instance, the cellular modem is by far the energy hog among the devices in WILDSCOPE. Nevertheless, our model shows that, when used at the rate typically demanded by biological studies, its aggregate contribution is marginal; instead, the configuration of GPS has the highest impact on lifetime, higher than proximity detection.

Contribution #3: Validation against ground truth. The other missing element is a *validation* of the accuracy of the platform. Proximity detection is based on low-power wireless, known to be highly dependent on, e.g., the particular device, the environment where it is used, the power setting. The distance at which proximity detection occurs and its reliability, both key to biological models, are consequently affected by all these factors. Nevertheless, comparison of detected contacts against ground truth is typically overlooked, thus potentially hindering the scientific inference based on this technology. This question has been largely ignored by both the biologging and WSN communities, since the two communities tend to disregard the limitations imposed by the technology, and the specific conditions of application, respectively. Interestingly, this happened also in the early times of GPS applications in wildlife studies [5]. Therefore, not only we provide biologists with a novel and powerful tool, but we also assess its quality w.r.t. ground truth. One could argue that a lot of knowledge exists about low-power wireless, e.g., as theoretical models and lab empirical evidence. These constitute a useful background but, alone, they cannot provide a direct answer. Instead, *in-field* empirical evidence is necessary, imitating the actual situations where the system is used, and taking into account the entire hw/sw artifact complete of the collar casing.

The performance of WILDSCOPE is analyzed in Section 6, based on three experimental settings where we retain decreasing degrees of control but always provide ground truth. First, we report about "in vitro" highly controlled outdoor experiments without animals, where we study the relationship between proximity detection and the distance at which it occurs, in relation to factors known to affect wireless communication, i.e., radio transmission power, distance from ground, node casing. This is also the opportunity to compare against a commercial, state-of-the-art device, showing the peculiarity of each solution. The performance of proximity detection is then studied "in vivo", in an in-field setting where collars are deployed on horses and ground truth comes from direct observation by biologists on our team. To our knowledge, this is the first study providing quantitative evidence about the different contact rates between controlled and real conditions. Last, we analyze the operation of the whole system based on a 2-month dataset from a deployment where a free-ranging roe deer is monitored by WILDSCOPE. Unlike other studies, we report about its accuracy w.r.t. ground truth, obtained via strategically-placed camera traps. Overall, this empirical evidence confirms the reliable operation of WILDSCOPE and provides guidelines for its use in ecological studies.

Finally, Section 7 ends the paper with brief concluding remarks.

2. RELATED WORK

Biologists have used UHF proximity loggers (mainly from Sirtrack Ltd.) to study individual encounters for almost a decade [10], to investigate a wide range of wildlife ecology issues, especially related with disease transmission (e.g., [3]).

The WSN community recognized wildlife monitoring as an application domain early on. DuckIsland [17] described a static WSN for monitoring storm petrels. Zebranet [16] was arguably the first project to employ animal-borne WSN technology, exploited to wirelessly transmit the acquired GPS data opportunistically until a base station is reached. Recent years have seen renewed interest in the topic. Dyo et al. [8] describe a system to monitor European badgers using animal-borne RFID tags. WSN nodes are only fixed, used for environmental monitoring and as communication relays. CraneTracker [1] is designed to monitor the location of whooping cranes via GPS, whose data is remotely transmitted via a cellular modem. Camazotz [11] provides a sophisticated platform targeting flying bats, where the localization schedule of GPS is determined by activity sensors, to optimize its duty cycle. Neither system provides proximity detection; low-power wireless is used, but only as a means for data offloading or to enable operator access to the nodes.

None of these works provide the integrated functionality of georeferenced proximity detection we introduce. The novel option to trigger GPS acquisitions upon contacts directly ties together location and proximity information. As for data offloading, we provide both options of in-field collection via fixed low-power wireless nodes and remotely via cellular modem. Finally, all of these works focus on a given species and for the purpose of a given study; extensions to other species is rarely mentioned. On the contrary, we enable application to multiple species and biological questions with a modular hardware and collar design, but especially through our validation results and the lifetime model, which together enable biologists to configure the platform to their specific research needs.

Although proximity loggers are gaining momentum, only very few studies (all in the biologists' community) assess their performance. Prange et al. [20] evaluates the effects of antenna orientation on contact distance for Sirtrack loggers, while Drewe et al. [6] includes also the effect of distance from ground and simulated body interference. Their experiments inspired our "in vitro" ones in Section 6 where we reproduce different animal positions (e.g., resting or eating vs. standing or moving). Moreover, as the environment and the animal body affect radio signal propagation, we also assess WildScope "in vivo" on animals, motivated by works [3,20] stating that these tests should be the rule, but this is seldom the case.

3. REQUIREMENTS

In this section we concisely outline the key requirements for WildScope established by the biologists on our team. The following define distinctive traits and the scope of our work:

R1. *Applicability to multiple species of different size.* Our original research interests were in ungulates (roe deer and red deer) and even larger animals like bears. However, biologists soon realized that studies of other smaller species such as foxes could also benefit from WildScope. The need for a single platform addressing animals of different species and sizes was therefore an early requirement.

R2. *Monitoring the use of focal points.* Recording the presence of an animal in a given place is also essential to ecological research; however, existing methods have major drawbacks. Camera traps have a limited view field, while the periodicity of GPS on collars due to battery constraints prevents continuous monitoring of the focal point. Therefore, biologists wanted to assess the opportunity offered by WildScope, whose low-power wireless proximity detection covers the entire 360° area around the node, and can be configured with a custom time resolution, as described next.

R3. *Focus on long-range interactions.* State-of-the-art proximity loggers focus on interactions occurring among individuals that are close (e.g., within a meter), motivated by initial studies on disease spreading. In studying the species above, we are motivated by different research goals, e.g., the above use of focal points or other animal social patterns, which entail the assessment of long-range interactions (e.g., several meters), for which existing proximity loggers are ill-suited.

Other requirements targeted geo-referenced proximity detection:

R4. *Custom time resolution for contact detection.* The *time resolution* of proximity detection is the minimum time interval to consider the co-presence of two individuals as biologically meaningful. A related parameter, used when proximity is no longer detected, is the *separation time*, i.e., the time interval after which a contact can be considered interrupted. Contacts shorter than the time resolution are ignored; contacts closed and reopened within the separation time are merged. The customization of these parameters is essential for biologists, who must adapt them to the species and studies at hand.

R5. *Periodic and contact-triggered GPS acquisition.* Acquiring the animal position upon and during proximity detection, the key functionality of WildScope, is very useful when contacts occur. However, it is equally important to continuously track animals to determine their complete trajectory and infer spatial use patterns. This motivates the parallel use of the typical, contact-agnostic, periodic GPS acquisition.

	mobile		fixed	
	DEER	FOX	ANCHOR	BASE
GPS	yes		no	
modem	yes	no	no	yes
contact detection	with all		with mobile	
offload data via radio	yes		to operator	
collect data via radio	no		from mobile	

Table 1: Functionality provided by WildScope nodes.

Several other lower-level requirements were considered, and are described in Section 4 hand-in-hand with the solutions we provide.

4. WILDSCOPE: PLATFORM AND TOOLS

WildScope mobile nodes come in two variants, shown side-by-side in Figure 1, targeting species of different size, as per requirement R1. We name these variants with the animal that motivated their development, i.e., DEER and FOX, respectively. The differences are determined essentially by form factor and weight. The guidelines commonly adopted for bio-logging studies state that devices must not hamper the comfort or otherwise cause harm to the animal (e.g., the node and battery casing should not be too large, sharp, or with elements that can get caught in vegetation) and limit the total weight (i.e., collar included) to 8-10% of the weight of the host animal. DEER measures 6.4×3.9×2 cm and weighs 34 g, while FOX measures 4×2.8×1.2 cm and weighs 14 g. The overall weight with battery and collar is 440 g and 240 g, respectively.

To fulfill requirement R2, WildScope employs fixed nodes deployed in the animals' habitat, which also come in two variants called ANCHOR and BASE. Both can be used as landmarks, to record proximity to mobile nodes and measure the use of a focal habitat resource; BASE also provides the ability to remotely transmit data via modem. The components installed, and functionality provided, of the various types of nodes are summarized in Table 1.

4.1 Hardware

Mobile nodes: DEER and FOX. The DEER mobile node, the first we designed and richest in features, is composed of two boards, shown in Figure 1. The *main board* contains the bare WSN node, whose design is similar to the popular TMote Sky, from which it differs in two respects. First, we use the TI MSP430F2618 MCU instead of MSP430F1611. The former provides a larger program memory (116 kB vs. 48 kB) and is better suited to our software architecture, which must manage many hardware components and their complex application functionality. The corresponding reduction in data memory (8 kB vs. 10 kB) and slower wake-up time (6 μs vs. 1 μs) do not pose problems in our case. Second, the main board uses a 2-Mbit FRAM (Ferromagnetic RAM) memory chip instead of the commonly used Flash memory. FRAM consumes less power than Flash, and offers faster write access and higher

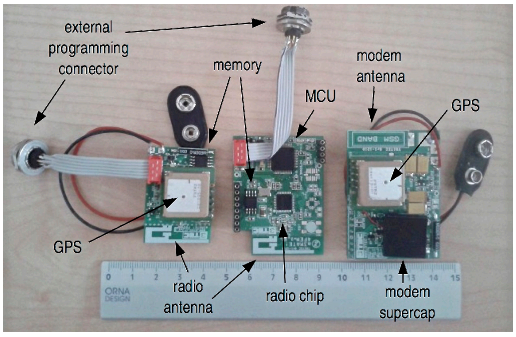


Figure 1: "Naked" mobile nodes, i.e., before packaging in a collar. Left to right: FOX, DEER main board, DEER child board.

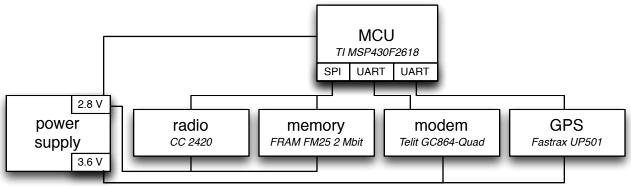


Figure 2: Architecture of the mobile WSN node.

limits on write-erase cycles. As in TMote Sky motes, our main board supports low-power wireless communication via a CC2420 transceiver and an on-board inverted-F microstrip antenna. FRAM and radio chip communicate with the MCU via the same SPI bus.

The *child board* contains the GPS and GSM modem chips, communicating with the MCU via separate UART buses. The GPS, a Fastrax UP501, is wired to a primary power supply, connected only when a localization (a "fix" in jargon) is requested, and an always-connected backup one, to preserve the data (e.g., ephemeris and satellite list) in the GPS RAM and minimize the time-to-first-fix upon activation. The modem, a Telit GC-864-Quad v2, is the most energy-hungry component. The battery alone cannot follow the modem abrupt energy consumption dynamics; a supercapacitor, expressly designed for mobile telephony, is therefore interposed between the two. Accommodating the modem on the board was also complicated by constraints on the relative position of its antenna connector and the FRAM. The electromagnetic field generated by the former caused corruption in the latter, and required a redesign of both the board and the firmware to ensure correct operation.

The primary power supply is a D-size Lithium battery that, with its weight of 110 g, is the heaviest component of the mobile node. The battery operates at 3.6 V, reduced to 2.8 V by a regulator on the main board, to reduce the consumption of MCU, radio, and FRAM.

The desire to target smaller animals forced us to reconsider the design of DEER, due to the limitations on form factor and weight. After consulting with the biologists, we concluded that the modem had to be sacrificed in FOX, being the component occupying most space on the board. This decision allowed us to optimize space by using a single board and to use a smaller 52 g C-size battery. The downside is that data offloading can no longer be performed remotely, and must rely entirely on the fixed nodes, described next.

Fixed nodes: ANCHOR and BASE. In WildScope, fixed nodes deployed in the habitat serve two purposes. First, they allow biologists to detect proximity of animals to landmarks (e.g., feeding stations, water ponds, or other relevant areas). As such, fixed nodes are part of the monitoring network as they *generate* contact data. Second, fixed nodes serve also as data sinks: when an animal-borne node is in range, the fixed node is able to *collect* data from it, as described in Section 4.4. The ANCHOR node hardware is not particularly interesting, as it has neither a GPS nor a modem; in principle, any node functionally equivalent to a TMote Sky will do.

BASE nodes provide another in-field outlet for gathering data. They behave just like ANCHOR ones w.r.t. contact detection, as shown in Table 1, but are additionally able to transmit the collected data via modem, just like mobile nodes. Actually, the hardware implementation *is* a mobile node, except the GPS chip is not mounted and an SD card provides larger data storage. Several power supply options are possible for BASE nodes; in our deployments, their long-term operation is ensured by a 12 V battery.

4.2 Collars

Deployed hardware must be protected by a proper packaging, which in wildlife monitoring must be attached (typically with a collar) to the animal, whose behavior is itself a threat to nodes (e.g., animals rolling or scratching against objects). The challenge is to balance the collar overall robustness and the animal comfort, not only because it may affect the animal's behavior—the subject of study—but also due to legal and ethical implications.

Our design aims to keep the two collar variants, shown in Figure 3, as similar as possible. DEER collars consist of two juxtaposed strips of plastic and rubber bent in a U-shape; FOX ones are made of a single leather strip; both have holes to adapt their length to the animal neck. The node is placed in a fiberglass resin box screwed to the collar top, and sitting between the shoulders where antennas (for radio, GPS, and modem) enjoy an optimal position. A second box at the bottom hosts the battery, whose weight stabilizes the collar. Cables protected inside the collar connect node and batteries.

Both boxes are filled with epoxy resin to protect their content from the external environment, which proved to be a crucial issue. In our first deployments, we used instead a simple ABS box, protected by a larger fiberglass one, and used straps similarly made with materials and techniques favoring rapid construction more than robustness. These decisions proved to be the *worst* we made in designing WildScope. The time saved in building collars was negligible compared to that wasted on chasing the source of faults. While we erroneously ascribed them to the immaturity of the hw/sw platform, faults were often induced by humidity and condensation in the boxes, cables unable to sustain the mechanical stress induced by animals, and other collar manufacturing problems. One practical lesson we learned is: do not compromise on collar design.

To enable reprogramming, we make the required pins available outside the node box through a metal strap wired to an IP67 (i.e., waterproof and dustproof) connector with screw terminals, further sealed with threadlocker glue before deployment. The node is also connected to a magnetic switch, the silver bulge at the top in Figure 3. This allows us to fully assemble the collar, yet turn it on right before deployment, saving battery and storage.

Both collars can accommodate a Lotek Wireless Inc. drop-off device. Visible on the left in Figure 3, it is designed to automatically open after a pre-defined, non-modifiable time, set during the manufacturing process. This device, commonly used in bio-logging studies, along with a means to locate collars in the wild (Section 4.5), allows node recovery after battery depletion.

4.3 Geo-referenced Proximity Detection

The main functionality provided by WildScope to biologists is the ability to detect proximity between animals using the low-power wireless radio as a "contact sensor", and simultaneously tag this information with the location where the contact occurred.

Proximity detection with low-power wireless. WildScope relies on a simple yet effective neighbor discovery protocol, illustrated in Figure 4. Time is discretized into logical *epochs* of duration E; epochs are *not* synchronized. Each node first sends a beacon of predefined duration b at the beginning of its current epoch.

Figure 3: Deployment-ready collars: DEER and FOX.

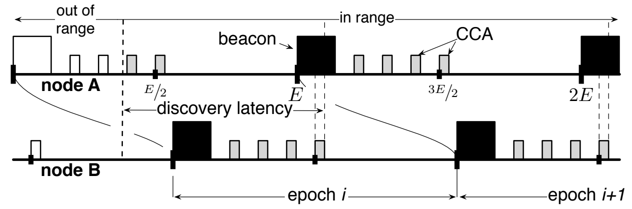


Figure 4: Neighbor discovery protocol.

Then, for the remaining first half $E/2 - b$, the node performs low-power listening (LPL) [19], i.e., it probes, with period $t_{LPL} = b$, for the neighbors' beacons via clear channel assessments (CCAs) and turns off the radio in between CCAs. The radio is turned off also for the *entire* duration $E/2$ of the second half of the epoch.

This scheme guarantees that at least one of the nodes is able to detect the other upon coming in range, regardless of the difference in the timeline origin between the two nodes. The maximum latency between the instant t_0 when the nodes become in range and the instant t_b when one of the two successfully receives the other's beacon is $t_b - t_0 \leq E$. Therefore, E defines the time resolution of contact detection set by biologists (requirement R4) based on the species and research question at hand.

This technique is asymmetric: only one nodes detects the other. We enforce symmetry by having the detector (B in Figure 4) send a unicast message to the detected node (A) upon receiving its beacon.

While other neighbor discovery protocols exist in the literature, there are several reasons we designed our own. The protocol is very simple, and its implementation mostly reuses already necessary components (e.g., LPL). Further, the epoch E provides biologists with a *single* parameter *directly* and *deterministically* defining *both* the time resolution and separation time of contact detection, and whose effect is much more intuitive to grasp, model, and use than, e.g, the configuration of primes and slots in [7,12]. Other protocols improving on detection speed [23] and lifetime [2] appeared after our initial deployments; given the effort of the latter and the system complexity, changing the key functionality of neighbor discovery was not a wise option, and improving detection speed beyond our deterministic bound is not a priority. Similarly, as shown in Section 5, the reasonably low duty-cycle of our protocol makes its energy consumption significantly smaller than the GPS one; further optimizations are likely to bring only marginal improvements.

Managing and storing contact events. The reception of the first beacon (or unicast message) from a neighbor causes the creation of a *contact tuple* $\langle ID_{self}, ID_{neighbor}, ts_{open}, ts_{closed}\rangle$ in the receiver's FRAM. The tuple contains the identifiers of the two nodes receiving and sending the beacon, along with timestamps recording the contact start and end. Timestamps are logical: the number of epochs elapsed since the node booted. A contact is considered *open* upon reception of the first beacon, and consequent creation of the contact tuple. Due to the periodicity of our neighbor discovery scheme, two nodes that remain in range keep exchanging messages. This allows us to define a contact *closed* when a predefined number m of beacons (the separation time) are missed. If m is too small, contacts may become fragmented due to spurious losses on the wireless link. If m is too big, separate contacts may be incorrectly seen as one, biasing the biological interpretation. In our deployments, the former is a lesser evil; hereafter, $m = 1$.

The opening and closing of a contact are stored as separate entries in FRAM; in the former case, a placeholder is used in place of ts_{closed}. A node may have multiple contacts open; for each neighbor, an open event is recorded only upon receiving the first beacon.

Dealing with fixed nodes. As shown in Table 1, fixed nodes detect contacts only with mobile nodes. This avoids that two or more fixed nodes in range continuously record each other's presence, filling their FRAM with worthless data. Therefore, the node type (mobile or fixed) is encoded as part of the beacon, along with the node identifier. Fixed nodes ignore beacons received from other fixed nodes, while mobile nodes process all beacons indistinctly.

Enriching contacts with location information. WILDSCOPE supports two types of GPS activation, as per requirement R5: periodic and triggered. The former is commonly adopted by bio-logging studies [4]; a fix is acquired with period t_{GPS}, typically on the order of hours, and independently of the animal behavior and context. Unlike available bio-logging platforms, WILDSCOPE also supports GPS activations triggered by proximity detection, monitoring the animal position upon and during a contact. Position is associated to a contact when it is opened. However, since biologists are also interested in monitoring the position *during* a contact, a fix is also acquired with period $t_{trigGPS} \ll t_{GPS}$ as long as the node remains in contact. In practice, geo-referenced proximity detection enables biologists to force a tighter schedule on GPS when contacts occur.

As GPS is energy-hungry we limit its activity in two ways. First, we observe that contacts may be fragmented due to radio packet loss, a phenomenon exacerbated when collars are deployed on social animals moving around the wireless range. Further, the GPS schedule associated to multiple contacts would unnecessarily increase the number of fixes. If fixes were taken blindly each time a contact is open, they would be too numerous—and of little biological significance. Therefore, a fix is valid for a configurable time interval $t_{noGPS} \leq t_{trigGPS}$; fixes scheduled or triggered within this interval are suppressed. Hereafter, we always use $t_{noGPS} = t_{trigGPS}$. Second, in areas where GPS reception is impaired, the receiver would try indefinitely to acquire a good fix, consuming energy. Therefore, we set a timeout of 3', after which the GPS is switched off and the failed localization event is recorded in FRAM.

Regardless of the activation mode, a successful GPS fix results in the storage of a 32 B tuple with location information; the mode is however also recorded as a flag. The tuple contains a subset of the location information associated with the NMEA sentence read from the GPS: physical time, latitude, longitude, altitude, HDOP (horizontal dilution of precision), number of satellites used, and 8 boolean flags denoting data quality. Moreover, the tuple contains also the epoch in which the fix has been acquired, enabling realignment with contact information, as discussed in Section 4.6.

The actual procedure to acquire a "good" fix required some in-field experimentation. When the GPS is active, it transmits one NMEA sentence per second, but we empirically determined that the first 10 sentences are often unreliable. Therefore, we wait until the 11^{th} sentence and then store the one with the minimum HDOP, or the first one below a given threshold indicated by the biologists.

4.4 Data Offloading: In-situ and Remote

Mobile nodes store data in FRAM, until they have an opportunity to offload it. Two modes are supported: in-situ and remote.

In-situ offloading allows a mobile node to transmit its data to a fixed node. On ANCHOR nodes, data is stored locally and can be later retrieved manually. On BASE nodes, data can be automatically transmitted via modem at designated times. In both cases, the fixed node coordinates transmissions when multiple mobile nodes are present, by determining their order and amount of data to transmit based on a linear combination of the node's leftover battery and number of stored records. The fixed node independently acknowledges each record received, which can be safely removed from the mobile one, freeing memory for new records. In-situ offloading relies on the correct placement of fixed nodes; for some species, e.g., foxes, this can be achieved with enough reliability.

Otherwise, remote offloading, enabled by the modem, can complement in-situ off-loading. For instance, deer can move widely, thus visiting infrequently the sites where fixed nodes are placed; on the other hand, the areas where they range are not always covered by cellular signal. Modem connections are always initiated by the mobile node on a given schedule, e.g., daily in our current deployments. These periodic connections also provide a keep-alive of sorts; the mere fact that the node establishes a connection is a proof that it is still functioning. For this reason, the transmission of the records in FRAM is always preceded by a concise summary of the node operation (battery level, amount and type of data present, and the last 5 GPS fixes) enabling operators to gain up-to-date information about the node status and the current area where it is situated.

4.5 In-field Access to Mobile Nodes

Operators occasionally need to locate nodes—not a trivial task when operating in the wild. Locating a collar is necessary when it must be removed, or when it has automatically detached from the animal due to the drop-off device. However, it is often needed also to verify that the node is still operational, or to enable wireless access to it, e.g., to download its data in the cases where the other means are not available (e.g., FOX collars do not have a modem).

Finding mobile nodes: (very) long-range. As is common in biologging studies, our collars include a VHF analog transmitter, enabling their detection over large areas—kilometers, depending on the operator's receiver directional antenna. The transmitter size is 2.6×1.9×0.5 cm, its weight is 7 g. In DEER collars, the device is powered by a small (19 g) dedicated AA Lithium battery, and is independent from the node hardware and its potential failures. The weight and size limitations of FOX collars, instead, require that the VHF device reuses the same battery as the node hardware. However, our VHF device emits a 35 ms pulse every 2 s, drawing an average current of 180 μA; its power consumption is therefore negligible w.r.t. the other components, as later shown in Table 4.

Finding mobile nodes: short-range. MOTEHUNTER [13], developed in our group, relies on the low-power wireless radio to enable short-range detection. The tool operates on IEEE 802.15.4 packets; a "ping" packet with the A flag (ACK request) set is sent in broadcast, forcing a reply from nodes in range. MOTEHUNTER can be used without dedicated firmware components; we used it to locate nodes we had already deployed before completing the software. Nevertheless, a small component can be integrated on the mobile nodes which provides additional features, e.g., disabling LPL and increasing the transmit power upon receiving the ping message.

Interaction with mobile nodes. This functionality is enabled by using an ANCHOR node carried by the operator and attached to a PC. The firmware loaded on the ANCHOR replaces the automatic coordination of data offloading from multiple nodes with a manual one, driven by a GUI available to the operator through the PC. The GUI visualizes the nodes in range, and allows the operator to determine if and when data from a given node is to be downloaded.

4.6 Using the Data

The last component of the WILDSCOPE toolset is the database where the data gathered is stored. We use PostgreSQL 9.2 with the PostGIS 2.1 extension enabling data geo-referencing. The data transmitted by the mobile node via modem are automatically stored in the DBMS; those downloaded in-situ by operators are instead stored via custom scripts. Beside the mundane purpose of providing a single and well-structured repository, nonetheless fundamental for biologists,our database implementation provides two important functions: timestamp realignment and duplicate removal.

Timestamp realignment. The data recorded by nodes are timestamped with the logical time (epoch) of acquisition. However, the logical time is synchronized neither across nodes nor w.r.t. global (physical) time. The stored procedures in the database perform the appropriate conversion by using as a baseline the last reset (e.g., normally, when the node has bootstrapped) combined with a GPS fix, to provide a simultaneous reference for both logical and global time. Resets may happen due to external causes; for instance, our initial deployments were severely affected by problems with the collar manufacturing that, as discussed in Section 4.2, caused an erratic behavior of the hardware. Our timestamp realignment procedure is in any case robust w.r.t. resets (which now occur very rarely), and deals automatically with the misalignments they may induce. To account for clock drifts, logical and global time are kept synchronized through the GPS readings. Fixed nodes, without on-board GPS, synchronize via contacts with mobile or operator nodes. Clock drift is in any case not a dramatic issue, at least in our current deployments where we use $E = 60$ s as the epoch length.

Duplicate removal. Duplicates can arise for two reasons. The first is that we designed the data transfer mechanisms described in Section 4.4 to minimize data loss. For instance, if the transmission of a record is not acknowledged, the latter is not removed from the memory. If the record has actually been received, and the record is re-sent at a later time, a duplicate is created. The second reason is the fact that data may reach the database through separate paths—in-situ and remote. Imagine a record $\langle A, data, t\rangle$ is present in the database, and an identical record (a duplicate) is inserted. It is possible that, at a later point, data collected in-situ is inserted in the database; these data may contain reset information causing an automatic timestamp realignment. As a consequence, the duplicate may be changed to $\langle A, data, t'\rangle$ and no longer be such (or become a duplicate of a *different* record). For this reason, the database automatically marks records as "original" or "copy", but never removes the latter, as they can later change their status. Moreover, this choice doubles as a way for system designers to assess the actual number of duplicates generated by data offloading mechanisms.

5. CONFIGURATION: LIFETIME

Threats to lifetime come from two issues. Reliance on the limited energy budget of the battery eventually determines the inability to operate when this is depleted. A more subtle threat is posed by the data memory: if the mobile node is unable to offload the data collected, either in-situ or remotely, the memory eventually fills up and the node is unable to record additional data.

Whatever the threat, maximizing lifetime is not a simple task, as it requires a deep understanding of the platform innards. This section addresses precisely this need, and provides biologists with a model of how various (system and biological) parameters affect lifetime. The model, focusing first on energy and then data storage, is useful also to evaluate hardware changes, e.g., the gain achievable by replacing the GPS with a newer and less energy-hungry one, or using a memory with a different capacity.

Table 2 summarizes the model parameters. Some are characteristic of our hardware (e.g., the current draw of the various devices) or, like battery and memory capacity, depend on its specific variant (i.e., DEER vs. FOX). Others depend on the specific configuration of the firmware used. Hereafter, we use as a reference the parameters used in the roe deer deployment we describe in Section 6.2. Nevertheless, measuring some of the low-level quantities using real parameters would be impractical, as in the case of the current draw generated by the *daily* activation of the modem. Therefore, Table 2 also reports the configuration we used in our lab measurements.

	Parameter	Value
Platform	average current draw for device x, i_x	see Table 4
	duration of CCA, t_{CCA}	15 ms
	FRAM capacity	2 Mbit
	battery capacity, B	13 Ah
Configuration (deployment)	epoch, E	60 s
	LPL period, t_{LPL}	1 s
	GPS period outside contacts, t_{GPS}	3 hours
	GPS period inside contacts, $t_{trigGPS}$	15’
	min. interval between fixes, t_{noGPS}	15’
	GPS timeout	3’
	Modem period, T_{modem}	24 hours
Configuration (measurements)	GPS period	10’
	Modem period	10’
	Data records sent by modem	200
	Radio power setting	3

Table 2: Parameters of the lifetime model.

5.1 Energy

Measurement setup and results. Measurements of current draw were performed with an Agilent 34411A multimeter on a DEER node powered by its battery; we used a partially used one, to have a more realistic (and conservative) estimate. The multimeter was configured with a sampling period of 40 ms, a range of 1 A, and a resolution of 0.2 μA. The sampling period is a compromise between the precision required to distinguish the contribution to current draw of the various devices, and a duration long enough to observe the combination of multiple device activations. Overall, we collected 1 million samples for a total duration of 11 hours.

These samples have been classified automatically based on their value and sampling time, mapping each of them to a given “event”, i.e., the activation of a given device; Table 3 shows some statistics, while Figure 5 shows an example containing activations of different devices. Note for instance the peak of power consumption (271.6 mA) when the modem is turned on, as mentioned in Section 4.1. It should be noted that our measurements do not separate the contributions of multiple devices. For instance, acquiring a fix entails MCU computation and FRAM access. Further, GPS activation is likely to overlap with radio activation. All of these are lumped together in a single sample. This is actually a faithful representation of reality where these contributions do occur simultaneously, and would be hard to separate them in a model. Given the high number of samples we collect, however, we argue that our results about average consumption are valid for the estimates we derive. This reasoning holds in particular for the contribution we labelled as “background”. This contains everything except radio, GPS, or modem activation; it includes the consumption of MCU and memory, but also of on-board circuitry like the voltage regulator, the clock oscillator, and the GPS backup power.

Device	Activations	Average duration (s)	Total time active (minutes)
radio	17400	0.061	17.82
GPS	77	45.68	58.66
Modem	67	78.52	87.72

Table 3: Events observed during measurements.

	Samples	avg (mA)	stddev (mA)
i_{radio}	26744	10.5757	3.48
i_{gps}	87995	30.2886	3.61
i_{modem}	131587	93.4799	58.93
i_{bg}	753674	2.1717	1.43
Total	1000000	16.9197	37.59

Table 4: Average consumption for each type of device.

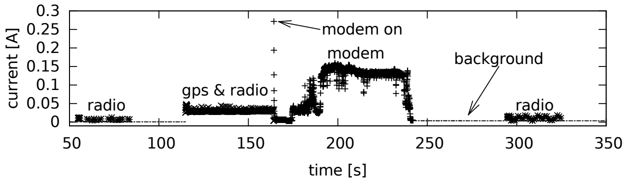


Figure 5: Current samples over a 5-minute interval.

Table 4 shows the measurement results; each sample is the average current draw observed during the 40 ms interval.

From measurements to lifetime estimates. Measurements provide the building blocks for a model of power consumption, whose equations are shown in Figure 6. Lifetime can be estimated by Eq. (1) as the ratio of battery capacity B and the sum of the instantaneous current draws I_x for the various devices x. The general form of the latter is given by Eq. (2), where i_x is the average consumption for device x in Table 4, T_{on} is the time interval where device x is active, and P the total functioning period under consideration. Next, we derive the actual estimates of I_x for each device.

Radio. If the node is alone for an entire epoch E, T_{on} is determined by the beacon transmission lasting t_{LPL}, plus the CCAs performed during the first half epoch, as shown in Eq. (3). Instead, if the node is in contact with N neighbors, Eq. (4) models the reception of their beacons and subsequent transmission of the unicast message, which on average lasts $\frac{t_{LPL}}{2}$. Beacon and unicast message have the same size of 61 B, yielding $t_{msg} = 1.95$ ms.

The parameters affecting consumption are the LPL period, t_{LPL}, and the epoch duration, E. However, t_{LPL} must be optimized for a given value of E, as the latter is set by biologists to determine the time resolution of contact detection. Nevertheless, the value of E is often not cast in stone and is determined empirically; knowing the impact of its value may help biologists to determine the best trade-off between biological value of the contact data and collar lifetime.

The relationship between current draw and t_{LPL}, for a given epoch, is represented by a curve like the one in Figure 7, where $E = 60$ s. In these curves, there is always an optimal value $\widehat{t_{LPL}}$; for $t_{LPL} < \widehat{t_{LPL}}$, the radio consumes too much energy in CCAs; for $t_{LPL} > \widehat{t_{LPL}}$ energy consumption is dominated by beacon transmission. For each epoch E we can easily compute $\widehat{t_{LPL}}$, i.e., the value that minimizes the current draw I_{radio} and therefore maximizes the lifetime L. The result is shown in Figure 8, assuming that the node is alone and no other device is activated.

GPS. The current consumption of the GPS chip is shown in Eq. (5), where t_{fix} is the average duration of a single GPS fix, and N_{fix} the number of fixes per day. Estimating t_{fix} is tricky, as it depends on the environment. If the animal is in a thick forest with limited sky view, satellite reception is hampered and t_{fix} may increase significantly, or reception may become altogether impossible. Therefore, we performed an in-field measurement campaign (using BASE nodes with GPS) in various areas, following the GPS acquisition procedures of Section 4.3, yielding $t_{fix} = 67$ s on average.

$$L = \frac{B}{\sum_x I_x} = \frac{B}{I_{radio} + I_{GPS} + I_{modem} + I_{bg}} \quad (1)$$

$$I_x = i_x \frac{T_{on}}{P} \quad (2)$$

$$I_{radio} = i_{radio} \frac{\left(t_{LPL} + \frac{E/2 - t_{LPL}}{t_{LPL}} t_{CCA}\right)}{E} \quad (3)$$

$$I_{radio}(N) = I_{radio} + i_{radio} \frac{(2t_{msg} + t_{LPL}/2)N}{E} \quad (4)$$

$$I_{gps} = i_{gps} \frac{N_{fix} t_{fix}}{1\ day} \quad (5)$$

$$I_{modem} = i_{modem} \frac{T_{conn}}{T_{modem}} \quad (6)$$

$$I_{bg} = i_{bg} \quad (7)$$

Figure 6: Energy model.

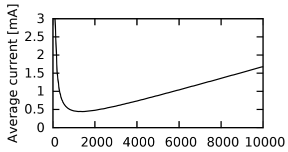


Figure 7: Current consumption vs. t_{LPL}, $E = 60$ s.

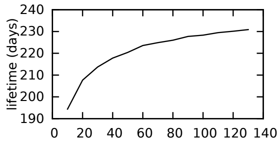


Figure 8: Lifetime vs. E (with the optimal $\hat{t_{LPL}}$).

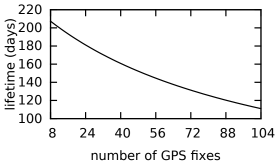


Figure 9: Lifetime vs. number of daily fixes, t_{fix}.

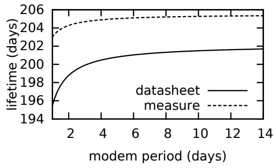


Figure 10: Lifetime vs. modem period T_{modem}.

The minimum value of N_{fix} is determined by periodic activations. In the configuration of Table 2, a GPS fix is acquired every 3 hours, corresponding to $N_{\mathit{fix}} = 8$ daily activations. However, additional ones can be triggered by contact detections, whose number is not known in advance but is limited by the fact that two GPS activations must be spaced apart by $t_{trigGPS} = t_{noGPS} = 15'$ in our configuration, yielding a maximum $N_{\mathit{fix}} = 104$. Figure 9 shows the estimate for N_{fix} ranging between these two extremes, assuming the radio is configured with the parameters in Table 2.

Modem. The consumption of the modem is shown in Eq. (6) where T_{conn} is the connection duration, and T_{modem} the period with which the modem is activated. The former can be further decomposed as $T_{conn} = t_{setup} + t_{data}$, where t_{setup} includes the time for modem initialization, acquiring a network, and setup and tear down the TCP connection, and t_{data} is the time spent only in data transfer. We verified experimentally that, in the areas with scarce coverage where our target species dwell, the total time to setup a connection is, on average, $t_{setup} = 30$ s. Similarly, by examining the real data from our deployments we established that, on average, a daily modem connection transfers 1105 B in $t_{data} = 21.40$ s.

Figure 10 shows the lifetime computed, as in the other cases, by using the deployment parameters in Table 2 and with an interval between modem activations, T_{modem}, varying between 1 and 14 days. In the chart, we take into account the fact that a higher value of T_{modem} implies a longer connection, as data from a higher number of days must be transmitted. This is done by simply considering $T_{conn} = t_{setup} + t_{data}T_{modem}$. The chart also considers two current draw values. The first, $i_{modem} = 93.4$ mA, is the one in Table 2, measured in the lab with good coverage. Coverage is likely to be worse in the wild, determining a higher consumption, which is however difficult to determine precisely. Therefore, we also use the (average) value $i_{modem} = 264$ mA in the datasheet, considerably higher than the first. In both cases, the curve raises steeply initially, but the overall difference is not significant. In the worst case, varying T_{modem} from 1 day to 2 weeks increases the lifetime by less than 1%. Therefore, although the modem is the most energy-hungry component, its impact on consumption in the long term is dwarfed by those of the other components, used more often.

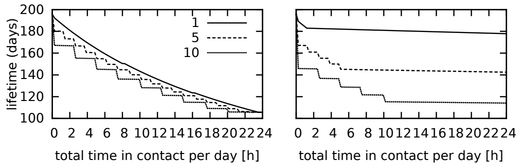
Background consumption. The last contribution comes from the background consumption present when none of the previous devices (radio, GPS, modem) is active. We account for it in Eq. (7) by approximating it with the average consumption in Table 4. This effectively *over*estimates consumption (and therefore *under*estimates lifetime) as it adds a constant contribution even when other devices are active. A finer-grained modeling would be very difficult, and the difference not very significant. We verified this last statement on our measurements, where we know precisely when the various devices are activated. If the background current i_{bg} is added to each of the 246,326 non-background samples, the average total consumption raises from 16.19 to 17.45 mA, a 7.78% increase.

Putting it all together. The last tiles to the puzzle of estimating lifetime are the number N_c of daily contacts and their average duration t_c. These bear a significant effect on lifetime (and data storage, discussed later), as they affect the active time of both radio and GPS. Unfortunately, these two parameters are often precisely the biological unknowns WILDSCOPE helps discover. To provide a frame of reference, analysis of the dataset in Section 6.2 shows that the DEER collar detected 385 contacts with the ANCHOR on the feeding station, averaging $N_c = 5.74$ contacts/day and $t_c = 9.96'$. These triggered 30.9 GPS fixes/day, in addition to periodic ones.

What really matters w.r.t. energy consumption is the *total daily time in contact*, T_c, affecting directly the active time T_{on} of the various devices. Even with a model of contact number and duration, contacts with different neighbors could overlap in infinite and arbitrary combinations, yielding significantly different T_c. Therefore, we base our estimate on a slightly different modeling of contacts, "fusing" overlapping contacts into a single one; two contacts $c_a(t_1, t_2)$ and $c_b(t_3, t_4)$ $t_2 > t_3$, are considered a single contact $c(t_1, t_4)$. This approximation is valid because *i)* energy consumption is dominated by GPS activation, whose energy consumption is about 3 times higher than the radio (Table 4), and *ii)* when a node participates in multiple contacts, and a minimum interval t_{noGPS} between fixes is enforced, a single fix is reported for all contacts; a single contact is equivalent to multiple overlapping ones w.r.t. GPS activations. We can define $N'_c \leq N_c$ as the number of non-overlapping contacts, and similarly $t'_c \geq t_c$. If a probabilistic contact model or real traces are available, based on N_c and t_c, it is easy to derive the corresponding values of N'_c and t'_c, which allow us to define the total daily time in contact simply as $T_c = t'_c \times N'_c$.

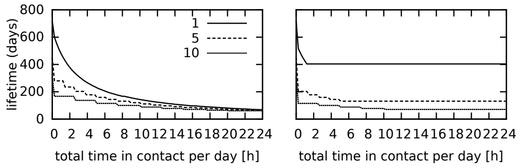
The last bit of information necessary to estimate lifetime is the policy governing the number of GPS fixes acquired while in contact. This is typically set by biologists based on the species and biological question under study. Here we consider two extremes. The NOLIMIT policy simply takes a fix as frequently as allowed by the t_{noGPS} parameter. The STARTEND policy, instead, takes N_{se} consecutive fixes, the first one when a contact is open, and similarly other N_{se} when it is closed, with the fixes spaced by $t_{trigGPS}$.

Figure 11 shows the lifetime estimate vs. T_c, for different values of non-overlapping contacts N'_c. The estimate considers all on-board devices (e.g., including modem activations) according to the deployment configuration in Table 2. Figure 11(a) uses the NO-



(a) NOLIMIT. (b) STARTEND.

Figure 11: Energy lifetime.



(a) NOLIMIT. (b) STARTEND.

Figure 12: Memory lifetime.

Limit policy, while Figure 11(b) uses startEnd with $N_{se} = 4$, as reasonable, e.g., for foxes. When noLimit is used, lifetime decreases steadily, and depends solely on T_c; the number of non-overlapping contacts N'_c does not bear a significant effect. The "steps" for $N'_c > 1$ are induced by the combined effect of t_{noGPS} over the non-overlapping contacts. When the startEnd policy is used, the average contact duration $t'_c = T_c/N'_c$ matters, as evident for $N'_c = 1$. If $t'_c < t_{noGPS}$ (15') the number of fixes is always N_{se}, as starts and ends are within the interval where a fix is "forbidden". If $t'_c > N_{se}t_{noGPS}$ (60'), the number of fixes is $2 \times N_{se}$, as the contact start and end are spaced apart by more than t_{noGPS}. In this case, the contact duration affects only the radio contribution, explaining the gentle slope for $T_c > 60'$. Therefore, this policy is convenient if a high T_c is expected, and detrimental if T_c is small.

5.2 Data Storage

The other component that can negatively impact lifetime is data storage. Our collars can offload data either to in-situ fixed nodes, or remotely via modem. However, for some species the optimal placement of fixed nodes may be difficult to guess. Moreover, the modem (not available on fox nodes anyway) can suffer from spotty cellular coverage rendering the animal isolated for long periods. Whatever the cause, if the node is unable to offload the data in FRAM, the latter eventually fills up; the node still functions, contacts and fixes are acquired, but they can no longer be stored.

At a minimum, a node stores daily 9 records; 8 (32 B each) for periodic GPS fixes and 1 about the outcome of the modem connection (26 B). The additional records stored depend, as for energy, on the number and duration of contacts N_c and t_c. The challenges to modeling are the same discussed earlier; we make the same approximations here, also for the sake of comparison, and determine lifetime as a function of the total daily time in contact, $T_c = t'_c \times N'_c$.

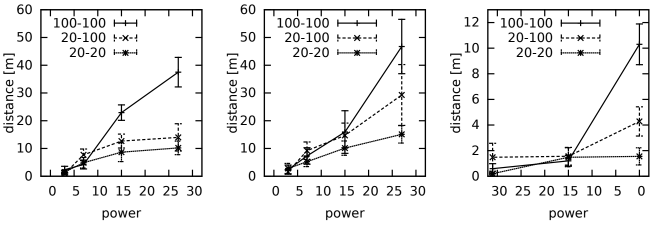
Figure 12 shows the results, taking into account the exact format of data records and the fact that the FRAM is also used to store system parameters. As with energy, lifetime depends on the GPS policy employed; again we compare noLimit and startEnd. The charts show that, with the configuration in Table 2, the bottleneck is usually energy consumption; data storage guarantees a lifetime higher than or, at worst, close to the one in Figure 11. Some configurations with $N'_c = 10$ are an exception: when using noLimit with $T_c > 300$' and startEnd with $T_c > 125$', data storage hampers lifetime before energy. Even in these cases, however, the memory lifetime is around 90 days. This aspect must be considered on a case by case basis—which is precisely the purpose of the model. In the worst-case scenario of insufficient memory lifetime, in-field data retrieval (e.g., by localizing the animal via VHF and deploying fixed nodes in the vicinity) is still a viable alternative.

6. VALIDATION: ACCURACY

In this section we validate WildScope along two dimensions. Section 6.1 focuses on the relationship between a detected contact and the distance between nodes, a fundamental parameter in ecological observations. Section 6.2 reports about an in-field deployment serving as a validation of the overall system.

6.1 Contact Detection vs. Distance

Here we focus on ascertaining the relation between contact detection and the distance at which it occurs in WildScope. We pursue this goal in two different ways. In Section 6.1.1 we perform experiments "in vitro", i.e., in a controlled and static setting, where we *measure distance given the possibility to control contacts.* Instead, in Section 6.1.2 we perform experiments "in vivo" with collared horses. In this setting, where we cannot control contacts, we *measure contacts given the possibility to observe distances.*



(a) Naked. (b) Coated. (c) Sirtrack.

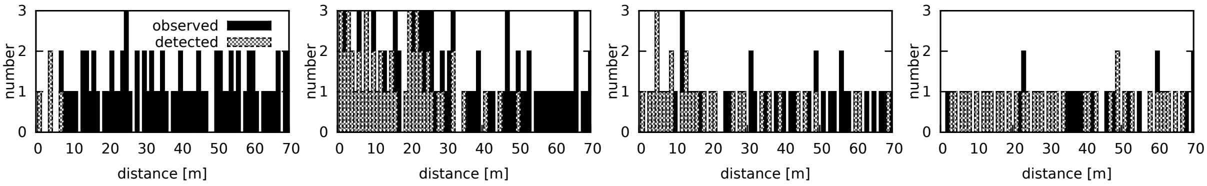
Figure 13: Effect of power and distance from ground on δ_{open}. Note the different scale on the y-axis for Sirtrack.

6.1.1 Experiments "in Vitro"

Goals and setup. The goal of these experiments is to determine the maximum distance δ_{open} at which contact detection begins. We use a controlled setup, whose design is adapted from [6, 20], in an open outdoor area without radio interference. Two mobile nodes (or loggers as the biologists call them) are placed on wooden easels, facing each other on a straight line. To imitate the effect of the animal body on wireless communication, we tied the loggers to a "neck" constituted by a 2-liter plastic bottle filled with a saline solution. The loggers are initially positioned far apart, where contact detection does not occur. Then, one of them is moved closer to the other, in steps of 0.5 m per minute. During each step, if the logger detects a contact, the current distance is recorded as δ_{open}, and the experiment terminated. We performed the experiments using 5 WildScope loggers and 5 Sirtrack commercial loggers. The latter provide only contact detection (i.e., no GPS) on the 915 MHz band, and are useful as a term of comparison. Both loggers are configured with a beaconing period (epoch) $E = 60$ s.

δ_{open} is affected by many elements, with radio *power setting* arguably bearing the most direct effect. We tested several powers on both loggers, from low to high (they use a different power numbering convention): 3, 7, 15, 27 for WildScope and 31, 15, 0 for Sirtrack. The settings for WildScope correspond to a nominal transmit power of -25, -15, -7, and -1 dBm, respectively. No information about the transmit power corresponding to a given setting is available for Sirtrack. Another element affecting δ_{open} is the casing, and in particular the resin coating of WildScope loggers. Assessing its impact on wireless communication, and therefore δ_{open}, allows us to relate experiments performed with naked nodes (common in the literature, easier to execute) vs. those made with deployment-ready nodes. Therefore, we considered both in our experiments. Finally, it is well-known that the *distance from ground* may affect wireless communication. Therefore, we performed experiments by placing the loggers at two different heights: 20 cm and 100 cm. This choice models a medium-sized animal (e.g., roe deer) resting on the ground and standing, respectively. In our experiments, we considered multiple combinations, denoted as 20-20, 20-100, 100-100, modeling animal interaction in different positions. For each experiment, in addition to δ_{open} we record the contacts *missed* (false negatives), i.e., those *never* opened, not even at zero distance between loggers. This metric is arguably even more important, as it directly impacts the reliability of the logger in *providing* a measurement. Each combination of logger, power setting, and distance from ground was repeated 20 times.

Results. Before we delve into our results, we comment about the reliability of contact detection. Sirtrack loggers exhibited a high number of missed contacts, all concentrated in 20-100 experiments, where they detected only 20% of contacts at low power 31, 35% at



(a) Power 3: 5% detected. (b) Power 7: 42% detected. (c) Power 11: 55% detected. (d) Power 15: 70% detected.

Figure 14: Contacts observed vs. detected as a function of power and distance for "in vivo" experiments.

intermediate power 15, and 100% only at high power 3. In contrast, WILDSCOPE coated collars *always* detected a contact in all our experiments. The fact that the Sirtrack false negatives occur in the 20-100 combination is probably a function of the antenna design and position, and its interaction with the radio chip: no information is publicly available about either. The finding is however very important, given that the 20-100 combination is interesting from an ecological perspective, as it models interaction between animals in different positions and therefore "states" (e.g., resting or eating on the ground vs. standing or moving). Further, these false negatives occur at low and intermediate powers, i.e., those commonly used in studies about disease spreading, as also reported in [6].

As for δ_{open}, its value increases with power for all loggers, as expected and shown graphically in Figure 13. However, somewhat to our surprise, WILDSCOPE coated nodes exhibit a greater δ_{open}, especially at high power. The resin coating and casing *increase* the transmission range, probably acting on the stability of the signal. Further, the trends for WILDSCOPE coated nodes in the 20-20 and 20-100 combinations are more linear than for naked ones; possibly, the resin helps reducing multipath interference from the ground.

We now focus on WILDSCOPE coated and Sirtrack, both used in real deployments. Table 5 offers a closer look at the data in Figure 13, and shows that distance from ground affects δ_{open} differently at different powers. Indeed, while for a given combination (i.e., column in Table 5) increasing power yields an increase in δ_{open}, this does not hold w.r.t. the distance from ground (i.e., across a row). We expected that, due to ground influence, δ_{open} increases when moving from 20-20 to 20-100 and 100-100. This trend is observed on both WILDSCOPE and Sirtrack, but only at intermediate and high powers: 15 and 27 for WILDSCOPE, 15 and 0 for Sirtrack. For lower powers, different trends hold. The lowest Sirtrack power shows a very small δ_{open} when nodes are at the same height (as small as 20 cm for the 20-20 case), and a much higher one for 20-100. The latter breaks the linear trend also in WILDSCOPE, but in different directions; the value for 20-100 is lower than the other two combinations at power 3, and higher at power 7.

Sirtrack loggers are more precise (i.e., less variance) than WILDSCOPE ones. This, combined with the ability to limit contacts to a very small δ_{open}, confirms that they are well-suited to observing close contacts (0.5–1 m), e.g., typical of disease transmission studies, although the strong presence of false negatives at 20-100 raises many doubts about the quality of the data gathered. In contrast, by fulfilling requirement R3 WILDSCOPE allows, for the first time, studies where contacts are defined on a larger scale, e.g., to assess the spatial interactions of medium-size animals (e.g., roe deer and foxes) with others or focal points. These interactions occur at several meters—a contact distance undetected by Sirtrack loggers.

WILDSCOPE coated			
power	**20-20**	**20-100**	**100-100**
3	280 ± 182	159 ± 71	238 ± 165
7	523 ± 175	920 ± 314	731 ± 297
15	1013 ± 258	1471 ± 444	1593 ± 768
27	1510 ± 314	2928 ± 1096	4675 ± 979

Sirtrack			
power	**20-20**	**20-100**	**100-100**
31	20 ± 8	148 ± 109	59 ± 39
15	148 ± 74	156 ± 69	120 ± 42
0	155 ± 67	428 ± 114	1030 ± 159

Table 5: Values (in cm) in Figure 13(b) and 13(c).

6.1.2 *Experiments "in Vivo"*

Goals and setup. In these experiments we measure the number of contacts detected by WILDSCOPE in a setting with animals, using as ground truth the observations of distances made by an operator. We placed 4 DEER loggers, complete with modified collar, on free-ranging horses in a fenced area of maximum length 150 m. This allowed us to test the effect of the animal body on contact detection. However, since animals were free to range, we could test the effect of neither distance from ground nor relative horse body positions on contact detection. We tested 4 power settings (3, 7, 11, 15) to confirm the relation with contact detection. We did not test power 27 as the tests in Section 6.1.1 confirmed that its range is too big for biological contacts. Each power was tested for 4 days.

An operator stood at the side of one animal, without interfering with its behavior and out of the line of sight between nodes, and measured its distance from other animals with a laser rangefinder. For each test session we performed several multi-minute trials where the distance between two nodes remained constant within a time interval between $E = 60$ s and 4'. If an animal moved during this interval, we interrupted the trial and waited at least 1' before starting a new one, to account for the separation time, $m = 1$. We recorded the start and end time of each trial, along with the pair of animals involved and the distance measured. Later, in the lab, we joined the distance data collected in-field with the contact data recorded by loggers based on their timestamp. We discretize time with a 1' granularity; for each minute and each (ordered) pair of loggers, we mark a 1 if a contact was detected, or 0 otherwise. We performed approximately the same number of trials for each power, covering all the distances in the 1–70 m range biologists deemed relevant.

Results. Figure 14 clearly evidences the impact of radio power on the number of contacts detected vs. observed, as a function of distance. These results are expected, coherently with the linear increase of power vs. distance discussed in Section 6.1.1. The very low number of recorded contacts for power 3, for instance, is due to the fact that, at that power, δ_{open} is about 2-3 m (Table 5), and horses were this close only on few occasions. However, in contrast with our in vitro experiments, we recorded false negatives even at very low distance and intermediate power, due to the mutual positions of the (massive) horse bodies, severely hampering radio communication. To our knowledge, this is the first time that the difference in contact rates between controlled and real conditions is assessed quantitatively. Biologist should take this stochasticity into account, e.g., by developing appropriate probabilistic models.

6.2 System-wide In-field Validation

Goals and setup. Next we moved away from controlled settings to validating the accuracy of the overall geo-referenced proximity detection in an in-field deployment consisting of a DEER collar, at-

Figure 15: Collar 44 and its host, during a visit to the feeding station. Behind the latter, a second, non-collared roe deer.

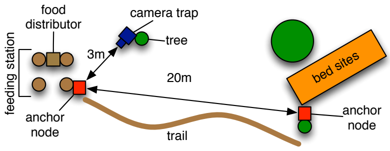


Figure 16: Sketch of the deployment area for Collar 44.

tached to a free-ranging roe deer, and two ANCHOR nodes, placed at a feeding station and bed site. Although small-scale, this deployment allows us to assess WILDSCOPE in the final conditions of operation, yet in a situation where we are able to provide ground truth. The deployment lasted several months, but only ~2 months, March 12 to May 17, 2014 could be used, due to a change of feeding habits and space use patterns of roe deer in late spring. Nevertheless, in this period we collected 1447 contacts and 1227 GPS fixes, as shown in Table 6. Only 48 fixes (3.9%, a very good ratio w.r.t. the state of the art) are invalid and excluded from analysis. Based on Section 6.1 and the biologists' interests, WILDSCOPE is configured with power 7. Other settings are reported in Table 2.

The DEER collar was deployed on March 6, on a male roe deer in an alpine environment. The animal was captured with a wooden box trap, placed at an artificial feeding station supplied with cereal pellets in a distributor accessible from three sides (Figure 15). The surroundings are a typical alpine mixed forest, with closed canopy and little understory vegetation, located on relatively steep slopes facing East (exposition= 50°), at 1108 m. Other individuals, some previously marked with eartags, access the same feeding station.

On March 12, we deployed an ANCHOR at height 1.5 m on a pole supporting the feeding station, ~1 m from the food distributor, and placed a second one 20 m from the feeding station, on a frequently used resting site (presence of bed sites). A camera trap was also placed at 3 m from the first ANCHOR, facing the food distributor. Figure 16 sketches the deployment. The camera, a Bushnell Outdoor HD Max 2012, is commonly used in wildlife studies. It is triggered by a passive infrared sensor, ensures minimum disturbance for animals, and enables color (day) or black-and-white (night) recording. We configured video mode with 60 s duration and 1 s time lap between triggers. Unfortunately, the camera trap proved less reliable than WILDSCOPE. It failed first on March 16; we discovered this and restored functionality only on April 5. The camera failed again on April 14, then we removed it. Therefore, our camera trap dataset covers ~15 days over two separate periods.

contacts	DEER collar		782	1447
	feeding station ANCHOR		383	
	bed site ANCHOR		282	
GPS (DEER collar)	invalid		48	1227
	valid	triggered	654	
		periodic	501	
		simultaneous	24	

Table 6: Data points collected during the deployment.

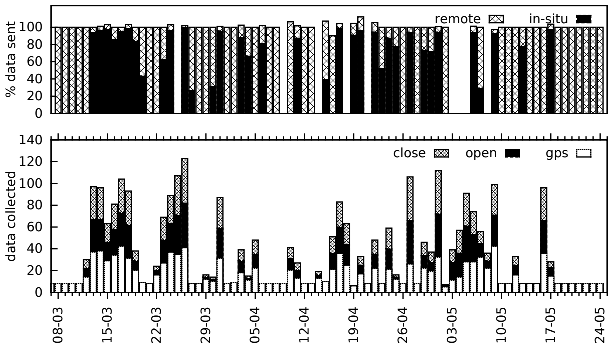
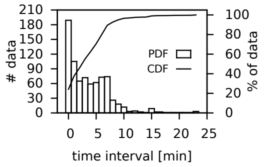


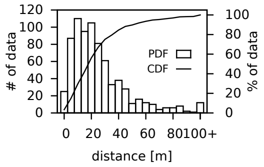
Figure 17: Daily modem reports (top) and statistics on the overall biological data (bottom) after manual in-field collection.

Results. Figure 17 shows information about the data gathered in the period under consideration. The top chart shows the daily reports from the modem, focusing on the percentage of data reported vs. collection mode, i.e., in-situ or remote. Modem transmission is reliable; complete data delivery is the norm and the impact of duplicates (values above 100%), discussed in Section 4.6, is limited. We occasionally lost contact with the node (e.g., first on March 25), likely because the animal was in an area with poor connectivity. Nevertheless, subsequent connections, in most cases on the day after, correctly transmit and report the records also for the missing periods, unless they have been already offloaded in-situ. The chart shows that in-situ offloading occurs for a significant fraction of the data, witnessing the effectiveness of our multi-modal approach.

At the end of June, we wirelessly downloaded the data in-field from the ANCHOR nodes as described in Section 4.5. After merging these data with the modem data in the database, we obtained the full dataset for the relevant period, also confirming the correctness of modem reports. The bottom chart in Figure 17 contains statistics about the *biological* data (GPS and contacts) gathered. The chart clearly shows when the animal is visiting the deployment area. Before March 12, the day we deployed ANCHOR nodes, only periodic GPS are reported, via modem. After that, the presence of contacts and triggered GPS tell us that the animal returned daily to the area, likely due to snowy weather. Regular visits stop on the first warm spring days around March 19, and resume a few days later. We speculate this relates to heavy snow falling the night before, pushing the animal towards the feeding station. These patterns repeat with different frequency, with the last recorded visit on May 17.



(a) Time distance.



(b) Spatial distance.

Figure 18: PDF and CDF for time and spatial distance.

We now analyze the quality of biological data, focusing first on position and its relation with proximity detection—the key novelty of WILDSCOPE—then on the performance of proximity detection w.r.t. the ground truth provided by the camera trap.

Figure 18 shows the probability (PDF) and cumulative (CDF) distribution functions of two important metrics. The first one is the time interval ΔT between the contact detection and the associated acquisition of the GPS location. Indeed, as we discussed in Section 5, in WILDSCOPE a triggered GPS may be "suppressed" if a fix has already been logged recently, i.e., within t_{noGPS}. Therefore, we define ΔT between the time of detection and the time of the *closest* GPS fix, regardless of whether it was periodic or triggered and before or after detection, consistently with the way biologists analyze the data. Figure 18(a) shows that in 89.3% of the cases, $\Delta T \leq {}^{t_{noGPS}}/_{2} = 7.5'$ and $\Delta T \leq 3'$ for 55.1% of the contacts. In practice, $\Delta T \ll t_{noGPS}$. Only 17 contacts (out of 782, i.e., 2.2%) have $\Delta T \geq 15'$. For 12 of these, $15' \leq \Delta T \leq 18'$, coherent with the 3' timeout after which the attempt to get a fix is aborted. As for the remaining 5 contacts (0.63%) for which $19' \leq \Delta T \leq 22'$, we are investigating this discrepancy, likely the effect of a rare corner case due to a timing interaction among devices that did not surface during earlier in-field tests. In general, this distribution of time distances is more than acceptable from a biological standpoint, compared to the state of the art in inferring contacts from GPS loggers.

The second metric is the spatial distance ΔS, computed between the GPS fix and the ANCHOR in contact. As shown in Figure 18(b), $\Delta S \leq 35$ m in 75.6% of the cases, which is in line with the biologists' expectations. The higher ΔS recorded in the other 24.3% can be ascribed to the animal movement during ΔT. Figure 19 plots the average and standard deviation of ΔS against ΔT. The chart shows that ΔS is under control for $\Delta T \leq 10'$, with outliers still compatible with the movement abilities of roe deer. For higher ΔT values, the likelihood that the animal has moved considerably from the detection point increases; moreover, the samples are too few (e.g., 1 contact for $\Delta T \in \{14, 18, 19, 22\}$) to be statistically significant.

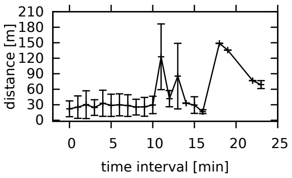


Figure 19: Time distance vs. spatial distance.

As for the performance of proximity detection vs. the ground truth of the camera trap, Table 7 compares WildScope contact data against visual inspection of the (timestamped) video frames. The first line accounts for contacts witnessed by the camera that WildScope was unable to detect either on the DEER or ANCHOR node. The occurrence of these false negatives is *extremely* low: only 2.04% of the contacts and 3.89% of the time detected by the camera are missed—significantly less than is reported in [15].

However, WildScope provides *significantly more* information than a camera trap. Focusing on contact duration, the most relevant to biologists, WildScope accounts for 97.8% of the total time detected, against 55.7% of the camera trap. Indeed, the latter is limited by the focal angle of the camera lens, while WildScope can detect the animal presence with a 360° angle. For biologists, assessing the real time an animal spends in a given spot (e.g., a feeding station) is of paramount importance; WildScope provides a significant advance of the state of the art in this respect.

7. CONCLUSIONS

We presented WildScope, a wildlife monitoring system that, for the first time, provides biologists with geo-referenced proximity detection. At its core, WildScope offers multiple hardware configurations adaptable to multiple species, parameterized software to enable the study of a wide range of biological queries, and a mathematical model to evaluate the impact of parameter choices on lifetime. We validated the system accuracy, offering biologists quantitative evidence of WildScope performance in real environments. Finally, we validated the platform on a real-world deployment on a roe deer, demonstrating the reliable operation of WildScope.

detected by	#contacts	duration
camera and not WildScope	1	0h 10'
WildScope and not camera	48	3h 23'
both WildScope and camera	29	4h 8' (WildScope) 4h 7' (camera)
camera (total)	30	4h 17'
at least one	78	7h 41'

Table 7: Contacts detected: WildScope vs. camera trap.

Acknowledgments. The authors thank Silvia Demetri, Noemi Giannini, Ştefan Gună, and Nathan Ranc for helping on this work.

8. REFERENCES

[1] D. Anthony et al. Sensing through the continent: Towards monitoring migratory birds using cellular sensor networks. In *Proc. of IPSN*, 2012.

[2] M. Bakht, M. Trower, and R. H. Kravets. Searchlight: Won't you be my neighbor? In *Proc. of MobiCom*, 2012.

[3] M. Böhm, M. Hutchings, and P. White. Contact networks in a wildlife-livestock host community: Identifying high-risk individuals in the transmission of bovine TB among badgers and cattle. *PLoS ONE*, 4(4), 2009.

[4] F. Cagnacci et al. Animal ecology meets GPS-based radiotelemetry: a perfect storm of opportunities and challenges. *Philosophical Trans. of The Royal Society Biological Sciences*, 365(1550), 2010.

[5] R. G. D'Eon and D. Delparte. Effects of radio-collar position and orientation on GPS radio-collar performance, and the implications of PDOP in data screening. *J. of Applied Ecology*, 42(2), 2005.

[6] J. Drewe et al. Performance of proximity loggers in recording intra- and inter-species interactions: A laboratory and field-based validation study. *PLoS ONE*, 7(6), 2012.

[7] P. Dutta and D. Culler. Practical asynchronous neighbor discovery and rendezvous for mobile sensing applications. In *Proc. of the 6th Int. Conf. on Embedded network sensor systems (SenSys)*, 2008.

[8] V. Dyo et al. Evolution and sustainability of a wildlife monitoring sensor network. In *Proc. of SenSys*, 2010.

[9] T. J. Habib et al. Modelling landscape effects on density-contact rate relationships of deer in eastern Alberta: implications for chronic wasting disease. *Ecological Modelling*, 222(15), 2011.

[10] W. Ji, P. White, and M. Clout. Contact rates between possums revealed by proximity data loggers. *J. of Applied Ecology*, 42, 2005.

[11] R. Jurdak et al. Camazotz: multimodal activity-based GPS sampling. In *Proc. of IPSN*, 2013.

[12] A. Kandhalu, K. Lakshmanan, and R. Ragunathan. U-Connect: A low-latency energy-efficient asynchronous neighbor discovery protocol. In *Proc. of IPSN*, 2010.

[13] C. Kiraly and G. P. Picco. Where's the Mote? Ask the MoteHunter! In *Proc. of SENSEAPP*, 2012.

[14] J. Krause, D. Croft, and R. Jamer. Social network theory in the behavioural sciences: potential applications. *Behavioural Ecology & Sociobiology*, 62:15–27, 2007.

[15] M. J. Lavelle et al. Assessing Risk of Disease Transmission: Direct Implications for an Indirect Science. *BioScience*, 64(6), 2014.

[16] T. Liu and M. Martonosi. Impala: A middleware system for managing autonomic, parallel sensor systems. In *Proc. of the 9^{th} Symp. on Principles and Practice of Parallel Programming*, 2003.

[17] A. Mainwaring et al. Wireless sensor networks for habitat monitoring. In *Proc. of 1^{st} Int. Wkshp. on Wireless Sensor Networks and Applications*, 2002.

[18] R. Nathan et al. A movement ecology paradigm for unifying organismal movement research. *Proc. Nat. Acad. Sci. USA*, 105(49), 2008.

[19] J. Polastre, J. Hill, and D. Culler. Versatile low power media access for wireless sensor networks. In *Proc. of SenSys*, 2004.

[20] S. Prange et al. New radiocollars for the detection of proximity among individuals. *Wildlife Society Bulletin*, 34, 2006.

[21] Y. Ropert-Coudert and R. P. Wilson. Trends and perspectives in animal-attached remote sensing. *Frontiers in Ecology and the Environment*, 3(8), 2005.

[22] E. M. Schauber, D. J. Storm, and C. K. Nielsen. Effects of Joint Space Use and Group Membership on Contact Rates Among White-Tailed Deer. *J. of Wildlife Management*, 71(1), 2007.

[23] D. Zhang et al. Acc: Generic On-demand Accelerations for Neighbor Discovery in Mobile Applications. In *Proc. of SenSys*, 2012.