A Collaborative Reputation Mechanism for Underwater Acoustic Networks

Filippo Donegà, Roberto Francescon, Filippo Campagnaro, Ivor Nissen, Michele Zorzi

***Abstract*—Distributed trust mechanisms can be used to protect Underwater Acoustic Networks (UANs) against a variety of Denial of Service (DoS) attacks. Several UAN trust systems have been proposed in the recent years, but only a few of them exploit Trust-Related Data (TRD) dissemination to improve the knowledge that a node has about its neighbors. Without TRD sharing, nodes can only form an opinion about their 1-hop neighbors, which might also be inaccurate due to packet losses being mistaken for actual malicious behavior. This paper presents a collaborative reputation algorithm for UANs that performs reputation sharing, propagation and aggregation allowing distant nodes in the network to form an opinion about each other and increasing the overall reliability of the trust metric. The proposed security mechanism is then implemented and tested as a DESERT Underwater module to verify its quality. Results show that trust values resulting from the aggregation of second-hand opinions are more reliable than the first-hand evaluations obtained through direct experience, effectively extending the knowledge that each node has about the trustworthiness of the others.**

***Index Terms*—underwater communications, underwater networks, security, trustworthiness, reputation.**

I. Introduction

Distributed reputation-based trust systems allow the detection of malicious activity within a wireless sensor network regardless of the specific threat. This is often more desirable than implementing specific countermeasures against each potential threat, given the quantity and variety of attacks that can be carried out against an inherently broadcast WSN deployment. Unfortunately, these mechanisms seldom suit more resource-constrained networks such as Underwater Acoustic Networks (UANs), where nodes have limited computational capabilities and the channel is characterized by low data rate, long latency and frequent link disruptions. Security in underwater networks has recently gained the interest of the scientific community [1]–[4] due to the increasing number of UAN deployments for mission-critical applications such as disaster monitoring [5], assisted navigation [6], mine detection [7] and many other strategic tasks [8], [9] that require high data integrity and timely fault detection and recovery. Terrestrial security mechanisms are generally incompatible with the constraints imposed by the underwater channel, such as the long propagation delay and the strong ambient noise, creating the need for further study in the field. Among the UAN trust systems proposed in the recent years [10]–[14], only some exploit Trust-Related Data (TRD) dissemination while satisfying the acoustic channel constraints [12]–[14].

The security mechanism proposed in this paper is built on a channel-based reputation system for UANs [14] that takes channel quality into account when evaluating nodes behavior to distinguish between actual malicious actions and packet losses due to poor channel state, judging nearby transmissions as compliant or non-compliant with the network protocol to build a first-hand reputation of 1-hop neighbors. On top of this, our module adds TRD sharing, propagation and aggregation, allowing a node to know whether or not all other nodes in the network can be trusted. Moreover, despite using [14] to gather one-hop trust evaluations in the context of this paper, our proposed algorithm is designed to be compatible with any other overhearing-based trust system capable of providing first-hand reputation assessments. An implementation of the proposed scheme is also provided as a module for DESERT Underwater [15], a publicly available [16] underwater network simulator developed and maintained by the SIGNET group at the University of Padova. Furthermore, the system is tested through DESERT simulations against a *dropping* attack to verify its accuracy and improvement on the baseline reputation system.

II. Collaborative Reputation Algorithm

The proposed security mechanism is based on TRD sharing, propagation and aggregation with first-hand reputations evaluated according to [14]. It consists of three main algorithmic components: first-hand TRD sharing, propagation of second-hand TRD and aggregation of second-hand TRD.

A. First-hand TRD sharing

Each node maintains an internal first-hand reputation table containing the trust scores of its one-hop neighbors, evaluated and dynamically updated according to [14]. First-hand reputation tables are periodically broadcasted by each node, and constitute second-hand reputation information (from now on, *recommendations*) for any node that receives them.

This work was partially supported by the Bundeswehr Technical Center for Ships and Naval Weapons, Maritime Technology and Research (WTD 71), Eckernförde, Germany, under the contract E/E71S/K1291/CF081; the FSE REACT EU, PON Research and Innovation 2014-2020 (DM 1062/2021); and the PRIN project BEASTIE (2022FZJ7MR), funded by the Italian Ministry of University and Research.

F. Donegà, R. Francescon, F. Campagnaro and M. Zorzi are with the Department of Information Engineering, University of Padova, 35131 Padova, Italy. F. Campagnaro and M. Zorzi are also with SubSeaPulse srl, 35131 Padova, Italy. M. Zorzi is also with Wireless and More srl, 35131 Padova, Italy (e-mail: filippo.donega@unipd.it, frances1@dei.unipd.it, campagn1@dei.unipd.it, zorzi@dei.unipd.it. Corresponding author: F. Campagnaro).

Ivor Nissen is with the Bundeswehr Technical Center for Ships and Naval Weapons, Maritime Technology and Research, D-24340 Eckernförde, Germany (e-mail: ivornissen@bundeswehr.org).

B. Propagation of second-hand TRD

Every node contributing to the reputation system rebroadcasts a received recommendation if all of the following conditions are satisfied:

1) the recommendation was never seen before;
2) the Time To Live (TTL) of the recommendation (i.e., a counter that keeps track of how many times the recommendation can still be re-propagated and is decremented at each hop) has not expired;
3) the latest recommendation forwarder is known and trusted according to the first-hand reputation table.

Conditions number 1 and 2 are enforced with the idea of minimizing the amount of data exchanged by the collaborative reputation system, in order to reduce the interference caused by security overhead. In fact, the proposed sharing mechanism does not aim at creating a common, coherent reputation agreement across the whole network, but instead focuses on expanding the knowledge that a node has about the rest of the network from few-hop neighbors to *n*-hop neighbors where *n* depends on the chosen TTL for recommendations. Condition 3 is introduced to counteract possible attacks aimed at the trust system such as *slandering*, where an attacker might propagate false reputation evaluations about some other node(s), or *replication attacks* where malicious nodes could replay recommendations to hide the fact that some other node's reputation has decreased or to fill the recommendation queues and delay updates. By filtering recommendations in this way, the opinion of nodes deemed unreliable is preemptively discarded.

C. Aggregation of second-hand TRD

Each node maintains a fixed-size vector for storing incoming recommendations and a second-hand reputation table that is periodically updated using the stored recommendations. The recommendation vector is managed as a FIFO queue: if there is no space left when a new recommendation is received, the oldest one is deleted and the new one is added. This design choice has the twofold purpose of saving local memory space and discarding old evaluations if more recent ones are available. Periodically, stored recommendations are aggregated, the resulting values are merged with the existing ones in the second-hand reputation table, and after this all recommendations are discarded. At every node, the aggregation process consists of the following steps:

1) for every node i, an aggregated reputation value $\texttt{aggr\_rep}_i$ is calculated as:

$$\texttt{aggr\_rep}_i = \frac{\sum_{j\in\Phi_i} w_j s_{ij}}{\sum_{j\in\Phi_i} w_j}, \tag{1}$$

where Φ_i is the set of nodes providing an opinion about node i, s_{ij} the opinion provided by node j about node i, and the weight w_j is the first-hand reputation value for node j;

2) if i is already present in the second-hand reputation table, its reputation value is updated as the weighted mean of the previously stored reputation $\texttt{old\_rep}_i$ and the new aggregated value $\texttt{aggr\_rep}_i$ with a configurable weight $\alpha \in [0,1]$ according to the following equation:

$$\texttt{new\_rep}_i = \alpha\,\texttt{old\_rep}_i + (1-\alpha)\,\texttt{aggr\_rep}_i\,, \tag{2}$$

If instead i is not listed in the second-hand reputation table, a new entry is inserted with reputation value equal to $\texttt{aggr\_rep}_i$.

At this stage, first-hand and second-hand tables are maintained separately, but depending on the application an integrated metric might be more desirable. In Section IV we will also evaluate the system's accuracy when using a weighted average of direct observations and external opinion.

III. Simulation Setup

The proposed security mechanism is implemented as a DESERT Underwater module called `UwSharedTrust` and tested against a simulated *dropping* attack. The network stack used in simulations is depicted in Fig. 1. Constant Bitrate (CBR) is used as Application layer to generate synthetic data at a constant rate. For Transport, a lightweight version of UDP with no checksum and 8-bit port addresses is employed. The `UwSharedTrust` layer is placed between the `UwFlooding` Network layer and the CSMA ALOHA Data Link layer. `UwFlooding` has direct access to the `NodeRep` object that implements the channel-based reputation system in combination with the `Overhearing` layer [14], and provides first-hand trust evaluations to `UwSharedTrust`. Finally, `UwPhysical` is used to simulate the physical layer according to the model in [17].

Fig. 2 depicts the grid-like topology used to test the system. The x and y coordinates of each node are randomly drawn within each cell according to a uniform distribution, while z (the depth) is constant and equal for all nodes. The node in the bottom-left cell is the network sink, while all the other nodes are sources generating packets of 40 Bytes every 300 s. All packets have the sink as their destination and propagate according to the `UwFlooding` Network protocol. The grid side

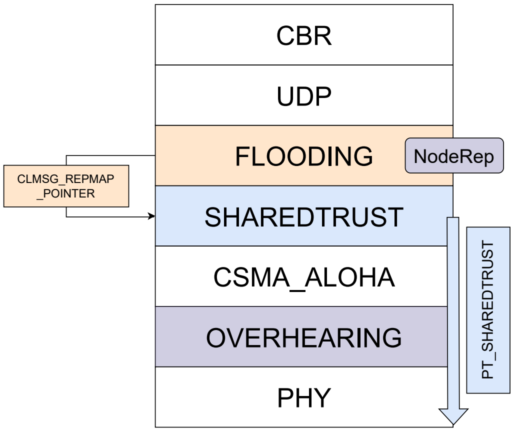

Fig. 1. Protocol stack used for simulations and main messages exchanged. `Overhearing` and `NodeRep` implement the channel-based reputation system [14], while `UwSharedTrust` handles reputation sharing, propagation and aggregation.

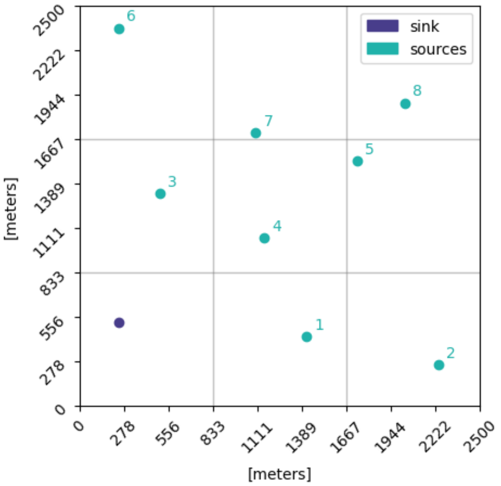

Fig. 2. 2D representation of a possible topology draw. In simulations, nodes are randomly deployed within each cell at every run. All nodes are deployed at the same depth of 1000 meters.

is set to 2500 m and guarantees good communication between nodes in horizontally and vertically adjacent cells. Instead, due to the random component of the topology displacement, nodes in diagonally adjacent cells might be out of range, requiring second-hand TRD to gauge the trustworthiness of one another. The TTL of both TRD and CBR-generated packets is set to 4, to avoid unnecessary re-transmissions while allowing some redundancy. `share_period` is arbitrarily set as twice the CBR period, while `aggregation_period` is twice the `share_period` so that more recommendations are stored before aggregation. A *dropping* attack is simulated by setting node 4 to start dropping packets with fixed probability `p_drop` at time 20000 s.

Nodes are modeled according to the EvoLogics S2C R 18/34 [18] Underwater Acoustic Modem, operating in the 18−34 kHz frequency band. To achieve a more realistic and meaningful simulation environment, ambient noise is introduced through wind and shipping activity models, as described in [17]. Specifically, the shipping factor is set to 1 (high activity), while the wind speed is set to 10 m/s (fresh/strong breeze) at the beginning of the simulation and varies every 10000 s according to a Gaussian distribution with $\mu = 10$ m/s and $\sigma = 1.2$ m/s. This results in a maximum transmission range that varies between 1.9 and 2 km. Interference is considered using the DESERT MEANPOWER model. The wave propagation geometry is set to circular due to the depth of the topology. Table I reports all values of the main fixed simulation parameters, excluding variable parameters such as simulation duration and packet drop probability which are tuned to fit specific analyses that will be discussed in Section IV.

IV. Results

Each result presented in this section is obtained by averaging over 25 runs to increase outcomes reliability, and reproducibility of results is ensured by using a seeded RNG. For the rest of this chapter, the term *second-hand table* will be used to shortly refer to the reputation table stored inside the `UwSharedTrust` module of each node, resulting from aggregation of recommendations, and periodically updated in the same fashion throughout the whole simulation. Furthermore, we will refer to all nodes different from the attacker as *legitimate nodes* for brevity.

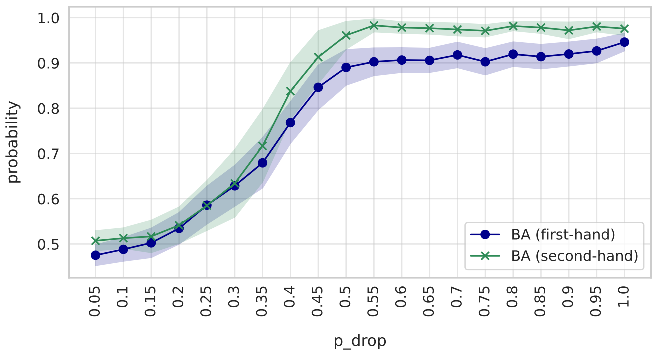

Fig. 3. Balanced Accuracy of the system for different values of `p_drop` at the end of the simulation obtained from first-hand (purple line) or second-hand (green line) reputation tables of legitimate nodes. Semi-transparent areas depict the 95% confidence intervals for the mean.

A. Accuracy test

This analysis evaluates the system's capability of reliably informing nodes about the presence of an attacker through TRD dissemination. 120000 s of deployment are simulated for different values of `p_drop` ranging from 0.05 to 1 with a step size of 0.05. Fig. 3 displays the Balanced Accuracy (BA) [19] of the system for each `p_drop`, obtained by averaging over all first-hand or second-hand reputation tables of legitimate nodes at the end of the simulation, and defined as:

$$BA = (TPR + TNR)/2, \tag{3}$$

where TPR is the True Positive Rate (i.e., the ratio of correctly classified malicious nodes with respect to the total number of nodes detected as malicious), and TNR is the True Negative Rate (i.e., the ratio of correctly classified legitimate nodes with respect to the total number of nodes detected as legitimate). The plot shows that second-hand reputation tables resulting from TRD aggregation are overall more accurate than first-hand tables populated through direct observation, probably

TABLE I
Fixed Simulation Parameters

Parameter	**Value**
Packet size	40 B
Bitrate	2500 bps
Transmission power	175 dB re µPa @ 1 m
Central frequency, bandwidth	26 kHz, 16 kHz
Shipping activity	1
Wind speed	Norm. dist. ($\mu = 10$ m/s, $\sigma = 1.2$ m/s)
Propagation geometry	Circular
TTL	4
CBR period	300 s

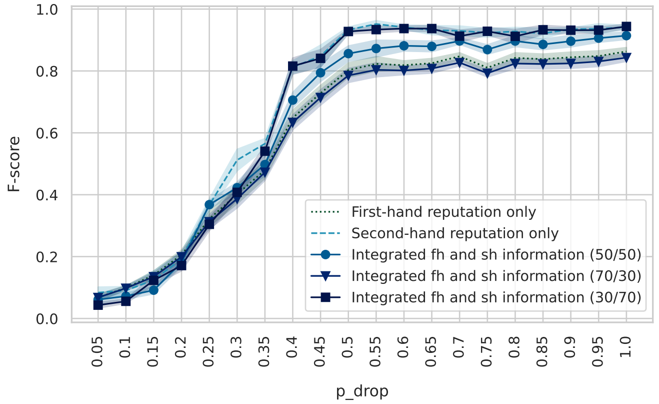

Fig. 4. F-score of the first-hand, second-hand and integrated first-hand and second-hand tables computed at the end of the simulation for different values of `p_drop`. Semi-transparent areas depict the 95% confidence intervals for the mean.

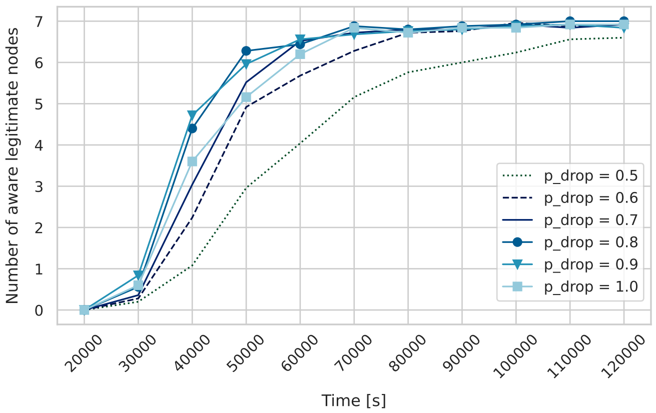

Fig. 5. Average number of legitimate nodes aware of the attacker thanks to second-hand TRD (i.e., classifying node 4 as malicious within their own second-hand reputation table) against simulation time for different dropping probabilities.

due to the fact that some first-hand false positives caused by particularly bad channel conditions get corrected by other nodes that manage to obtain a more reliable opinion thanks to a favorable channel. It can also be noticed that, in both cases, the system's accuracy becomes more stable for `p_drop` > 0.5, suggesting that, as can be intuitively expected, the channel-based reputation system has a harder time distinguishing actual malicious behavior from packet losses caused by a bad channel when the attacker drops a packet less than half the time.

The F-score of the system is also displayed in Fig. 4, computed as:

$$F\text{-}score = 2\,\frac{precision \cdot recall}{precision + recall}, \tag{4}$$

where *precision* and *recall* are the average precision and recall metrics of the following tables, extracted at the end of the simulation:

1) first-hand tables;
2) second-hand tables;
3) integrated first-hand and second-hand tables as a weighted mean with $\beta = 0.5$;
4) integrated first-hand and second-hand tables as a weighted mean with $\beta = 0.7$;
5) integrated first-hand and second-hand tables as a weighted mean with $\beta = 0.3$.

With integrated first-hand and second hand table of a node n we refer to an extension of the first-hand table T of node n that:

- includes the entries of the second-hand table S whose address a is not present in T;
- updates the reputation value rep_a of entries whose address a is present in S as a weighted mean with parameter $\beta \in [0, 1]$:

$$rep_a = (\beta)(t_a.rep) + (1 - \beta)(s_a.rep), \\ t_a \in T, s_a \in S. \tag{5}$$

Fig. 4 shows that for values of `p_drop` > 0.35 two alternatives present a better overall accuracy: using only second-hand reputations and using integrated first-hand and second-hand reputations with weight $\beta = 0.3$. This result further confirms the superior accuracy of second-hand evaluations, but also points out the 30/70 integration as a viable alternative that could be used to prevent some attacks (i.e., forging of recommendations) to which the system might be exposed when only considering second-hand tables for trust evaluation.

B. Dissemination test

Figure 5 displays the average number of legitimate nodes made aware of the attacker thanks to second-hand TRD against time, i.e., how many well-behaving nodes out of 7 list the malicious node as untrustworthy within their own second-hand reputation table. Different dropping probabilities are tested, ranging from 0.5 to 1 with a step size of 0.1, with node 4 as the attacker set to drop packets with constant `p_drop` starting at 20000 s. The plot highlights a direct correlation between the dropping probability and the number of legitimate nodes aware of the attacker at each time instant, while still demonstrating `UwSharedTrust`'s capability of making most (if not all) nodes agree on the trustworthiness of node 4 before the end of the simulation.

V. Conclusions and Future Work

TRD dissemination is proven beneficial to the reputation system, bringing significant improvements to the accuracy and scope of trust evaluations. In fact, second-hand tables resulting from aggregation are overall more accurate than the first-hand tables provided by the channel-based reputation system, and allow distant nodes to form a reliable opinion about one another. Furthermore, results suggest the proposed mechanism as a possible security solution for a scenario with mobile nodes, where nodes could keep track of the trustworthiness status of a mobile node that is not always within their range.

At the same time, it is important to notice that TRD sharing introduces some overhead whose impact needs to be analyzed, and exposes the system to attacks aimed at the reputation mechanism itself that will be addressed in our future work.

References

[1] C. Lal, R. Petroccia, K. Pelekanakis, M. Conti, and J. Alves, "Toward the development of secure underwater acoustic networks," *IEEE Journal of Oceanic Engineering*, vol. 42, no. 4, pp. 1075–1087, Jul. 2017.

[2] I. Ahmad, T. Rahman, A. Zeb, I. Khan, I. Ullah, H. Hamam, and O. Cheikhrouhou, "Analysis of security attacks and taxonomy in underwater wireless sensor networks," *Wireless Communications and Mobile Computing*, Dec. 2021.

[3] G. Yang, L. Dai, and Z. Wei, "Challenges, threats, security issues and new trends of underwater wireless sensor networks," *Sensors*, vol. 18, no. 11, p. 3907, Nov. 2018.

[4] R. Zhu, A. Boukerche, L. Long, and Q. Yang, "Design guidelines on trust management for underwater wireless sensor networks," *IEEE Communications Surveys & Tutorials*, vol. 14, no. 8, Aug. 2023.

[5] P. Kumar, P. Kumar, and P. Priyadarshini, "Underwater acoustic sensor network for early warning generation," in *IEEE Oceans*, Hampton Roads, VA, USA, Oct. 2012.

[6] M. Waldmeyer, H.-P. Tan, and W. K. Seah, "Multi-stage AUV-aided localization for underwater wireless sensor networks," in *IEEE Workshops of International Conference on Advanced Information Networking and Applications*, Biopolis, Singapore, Mar. 2011, pp. 908–913.

[7] S. Khaledi, H. Mann, J. Perkovich, and S. Zayed, "Design of an underwater mine detection system," in *IEEE Systems and Information Engineering Design Symposium (SIEDS)*, Apr. 2014.

[8] E. Felemban, F. K. Shaikh, U. M. Qureshi, A. A. Sheikh, and S. B. Qaisar, "Underwater sensor network applications: a comprehensive survey," *International Journal of Distributed Sensor Networks*, vol. 11, no. 11, Nov. 2015.

[9] M. Murad, A. A. Sheikh, M. A. Manzoor, E. Felemban, and S. Qaisar, "A survey on current underwater acoustic sensor network applications," *International Journal of Computer Theory and Engineering*, vol. 7, no. 1, Feb. 2015.

[10] Y. He, G. Han, J. Jiang, H. Wang, and M. Martinez-Garcia, "A trust update mechanism based on reinforcement learning in underwater acoustic sensor networks," *IEEE Transactions on Mobile Computing*, vol. 21, no. 3, pp. 811–821, Aug. 2020.

[11] J. Du, G. Han, C. Lin, and M. Martínez-García, "LTrust: an adaptive trust model based on LSTM for underwater acoustic sensor networks," *IEEE Transactions on Wireless Communications*, vol. 21, no. 9, pp. 7314–7328, Mar. 2022.

[12] J. Jiang, S. Hua, G. Han, A. Li, and C. Lin, "Controversy-adjudication-based trust management mechanism in the internet of underwater things," *IEEE Internet of Things Journal*, vol. 10, no. 3, pp. 2603–2614, Oct. 2022.

[13] G. Han, J. Jiang, L. Shu, and M. Guizani, "An attack-resistant trust model based on multidimensional trust metrics in underwater acoustic sensor network," *IEEE Transactions on Mobile Computing*, vol. 14, no. 12, pp. 2447–2459, Feb. 2015.

[14] A. Signori, F. Campagnaro, I. Nissen, and M. Zorzi, "Channel-based trust model for security in underwater acoustic networks," *IEEE Internet of Things Journal*, vol. 9, no. 20, pp. 20 479–20 491, Oct. 2022.

[15] F. Campagnaro, R. Francescon, F. Guerra, F. Favaro, P. Casari, R. Diamant, and M. Zorzi, "The DESERT underwater framework v2: Improved capabilities and extension tools," in *Proc. Ucomms*, Lerici, Italy, Sep. 2016.

[16] "DESERT: Design, simulate, emulate and realize test-beds for underwater network protocols," Last time accessed: Jul. 2024. [Online]. Available: https://desert-underwater.dei.unipd.it/

[17] M. Stojanovic, "On the relationship between capacity and distance in an underwater acoustic communication channel," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 11, no. 4, pp. 34–43, Oct. 2007.

[18] "EvoLogics S2C R 18/34D Underwater Acoustic Modem," Last time accessed: Jul. 2024. [Online]. Available: https://evologics.de/product/s2c-r-18-34d-27

[19] K. H. Brodersen, C. S. Ong, K. E. Stephan, and J. M. Buhmann, "The balanced accuracy and its posterior distribution," in *2010 20th International Conference on Pattern Recognition*. IEEE, 2010, pp. 3121–3124.