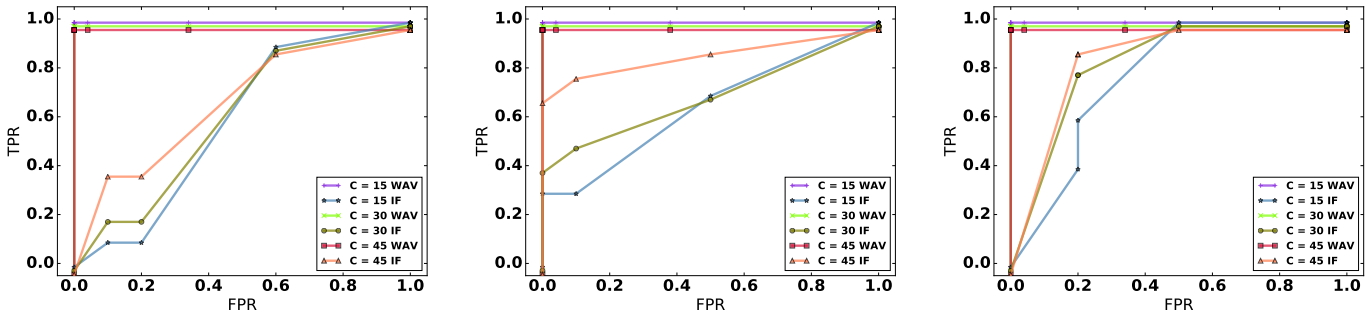
Supplemental Material for *Don't fool me!: Detection, Characterisation and Diagnosis of Spoofed and Masked Events in Wireless Sensor Networks*

Vittorio P. Illiano, Luis Muñoz-González and Emil C. Lupu

1 Performance with Increasing Number of Elicited and Masked Events

The performance of detection and characterisation of malicious data injections is affected not only by the number of genuine events but also by the number of eliciting/masking attempts. Below we present the experimental results in the wildfire monitoring WSN for an increasing number of elicited and masked events. In particular, we consider multiple eliciting scenarios where one, two, and three events are spoofed in turn starting from a rest condition. The eliciting detection results are shown in Fig. 1. We also consider masking scenarios, where three genuine events are present and one, two, and all of them are masked in turn. The masking detection results are shown in Fig. 2.

Note that the detectability of WAV does not show appreciable changes, since the ROC curve is nearly ideal in all scenarios. IF's performance instead has increasing performance with increasing number of elicited events. This is not surprising since the number of compromised sensors C is split into $\frac{C}{n_e}$ per event, with n_e being the number of elicited/masked event. Then, each manipulated event is less credible than a unique event manipulated by joining forces. However, with masked events there is an opposite relationship: more masked events are less detectable. This is because IF does not make any spatial reasoning about the measurements, but considers them as a generic collection. Hence, by removing the most event-like measurements from all events is better than focusing on the complete masking of a single event, since the measurements distribution becomes more homogeneous.



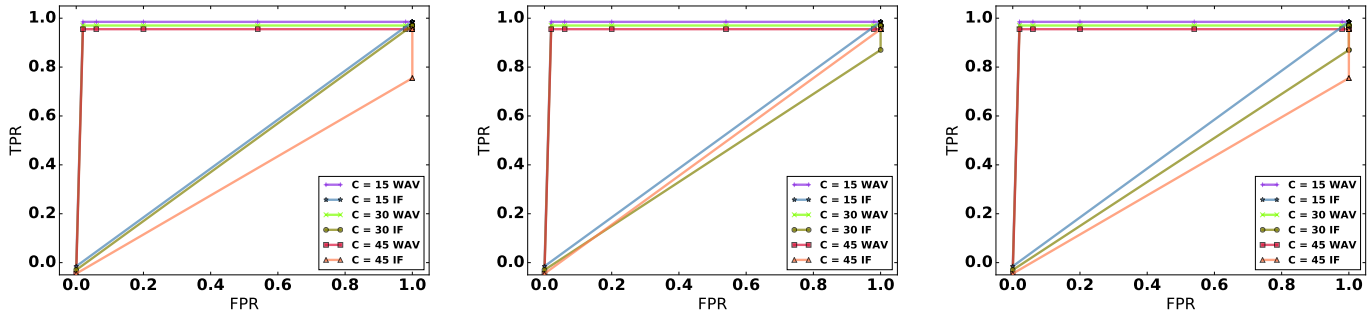
(a) 1 Elicited Event (b) 2 Elicited Events (c) 3 Elicited Events

Fig. 1: Eliciting Detection ROC curves with no genuine event VS number of elicited events. Performance increase due to a partitioning of malicious forces.

We present also the characterisation results of the same experiments: Fig. 3 and Fig. 4 show the performance with increasing number of elicited and masked events respectively.

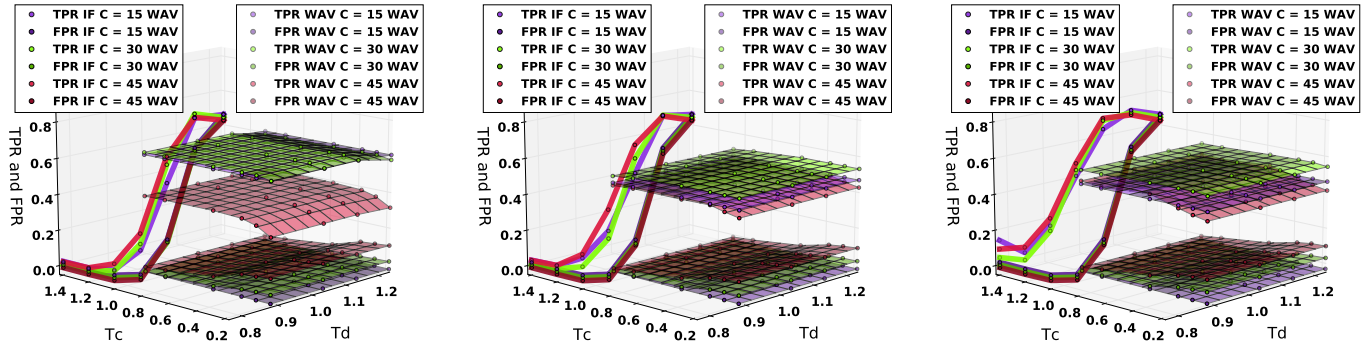
Note that IF shows a general increase in the characterisation performance with more elicited events and a decrease with more masked events, just like in the detection case. Instead, the performance of WAV depends highly on the number of malicious sensors. With increasing number of eliciting/masking attempts, the characterisation performance decreases with 15 malicious sensors and increases with 45 malicious sensors. The reason for the latter is the same as in the detection scenario: with more spoofed events, the collusion power is split. The worsening in performance with 15 malicious sensors instead is due to malicious sensors splitting into very small subsets, which creates noticeable gaps with the genuine rest measurements. As a consequence, the blending in sensors mediate more in favour of the genuine sensors and characterising

them becomes more difficult. On the other hand, such scenarios are still easily detected as anomalous, as the detection performance in Fig. 1 shows.



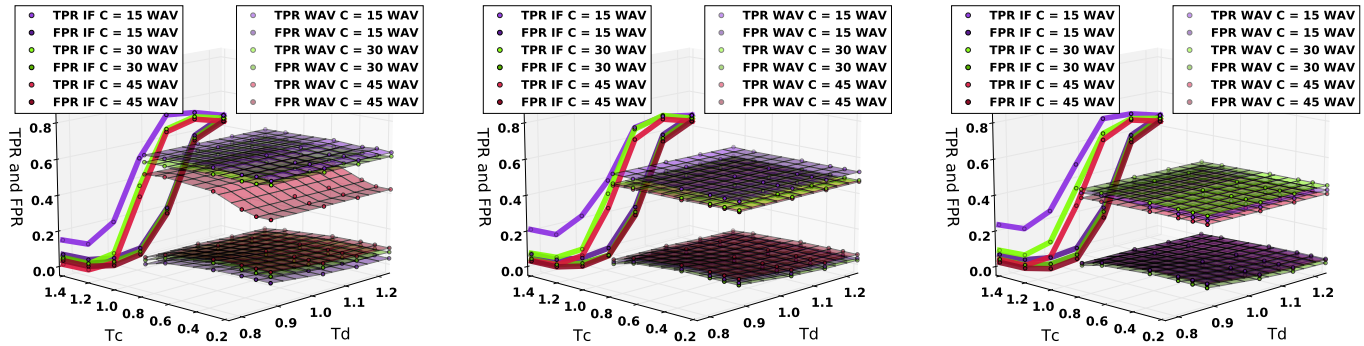
(a) 1 Masked Event (b) 2 Masked Events (c) 3 Masked Events

Fig. 2: Masking Detection ROC curves with 3 genuine events VS number of masked events. Performance increase due to a partitioning of malicious forces.



(a) 1 Elicited Event (b) 2 Elicited Events (c) 3 Elicited Events

Fig. 3: Eliciting Characterisation ROC curves with no genuine event VS number of elicited events.



(a) 1 Masked Event (b) 2 Masked Events (c) 3 Masked Events

Fig. 4: Masking Characterisation ROC curves with 3 genuine events VS number of masked events.

2 Performance with $r_i = 7$ in Wildfire Monitoring

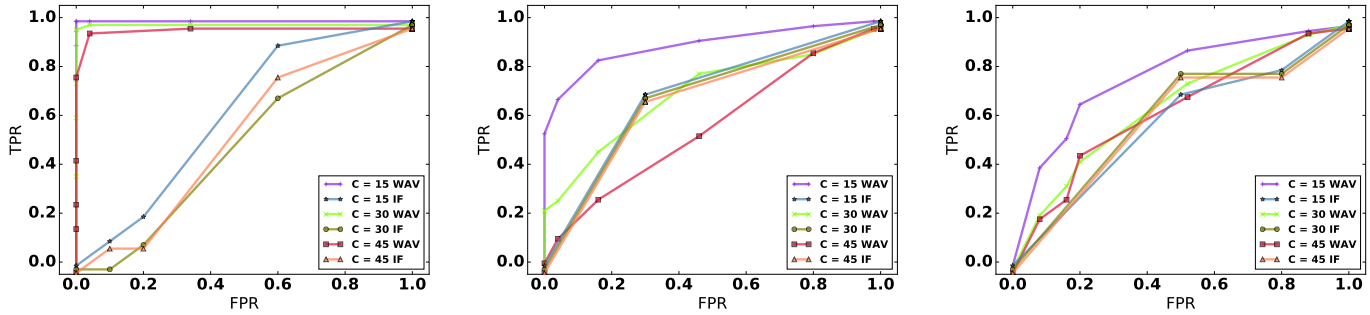
The events size is highly relevant to the performance of the algorithms against malicious data injections, especially when spoofed events are introduced or real events are masked. So far, we analysed the performance where the simulated wildfires have a radius $r_i = 22$ metres. This choice produces events that are perceived in about in 50% of the WSN deployment area (corresponding to 50 sensors on average), giving rise to perilous conditions since events are large enough to produce high damage when incorrectly detected.

Nevertheless, such large events require the attacker to compromise a large number of sensors, hence we also present the experiments when events have a smaller radius $r_i = 7$ and are therefore perceived in about in 15% of the WSN deployment area (corresponding to 15 sensors on average). Fig. 5 and Fig. 6 presents the relative detection ROC curves for elicited and masked wildfires respectively. Note that the detection is more complicated with smaller events, both for WAV an for IF. Indeed, spoofing or masking a smaller event requires less malicious sensors, since generally smaller events are perceived by less sensors. Nevertheless, the damage led to the WSN is also reduced assuming that the severity of the event is proportional to its size. If the latter is still high despite the reduced event size, then the detectability needs to be improved by increasing the density of the sensors deployment.

Comparing the two techniques, WAV still achieves better performance when the FPR is around 0.05. Moreover, WAV is capable of better distinguishing spoofed from genuine events as the number of malicious sensors decreases, while in IF such effect is not observed.

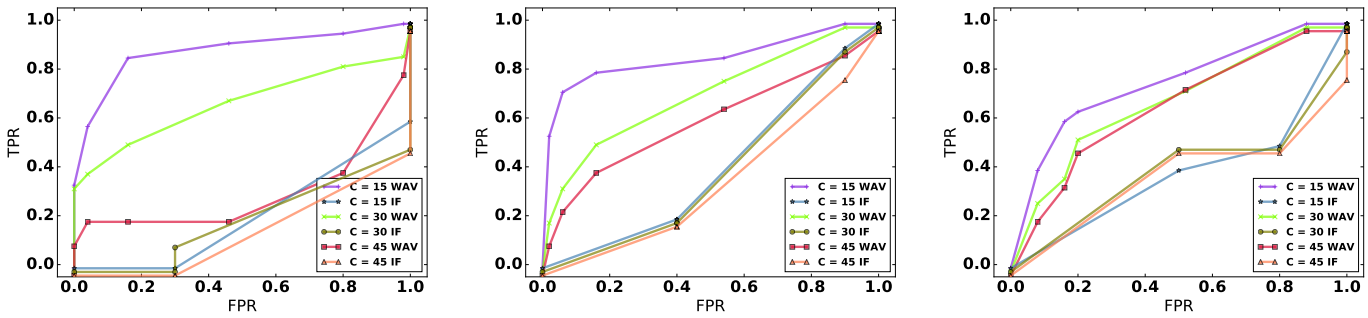
Note that, in comparison with the scenario where the radius is $r_i = 22$, here the detection performance of masked events does not generally increase with the number of genuine events. The reason for such result is that with smaller events it is more likely to observe areas experiencing a rest condition regardless of the number of events, while with multiple events with radius $r_i = 22$ there is a shortage of such measurements that introduces more conflicts between genuine event measurements and malicious rest measurements. With the support of genuine rest measurements, both when one and three events with radius $r_i = 7$ occur, the presence of more events can even become an advantage for the malicious measurements, since the non-masked events are a justification for the transitions between genuine event measurements and malicious masking measurements: in this case, a performance decrease with more genuine events can be observed.

Similar results hold for the characterisation task, whose performance is shown in Fig. 7 and Fig. 8. The closeness of the TPR and FPR curves in WAV, and especially IF, shows that the problem is complex in nature: identifying compromised events, perceived by an average of 15 sensors where 15 or more sensors can be malicious.



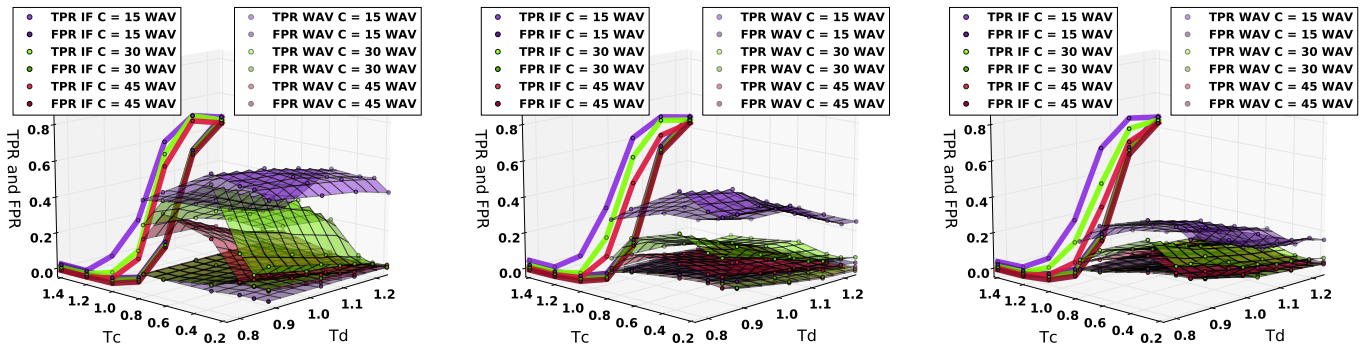
(a) No Event (b) 1 Event (c) 3 Events

Fig. 5: $r_i = 7$: Eliciting Detection ROC curves VS number of events. Performance decrease with more events.



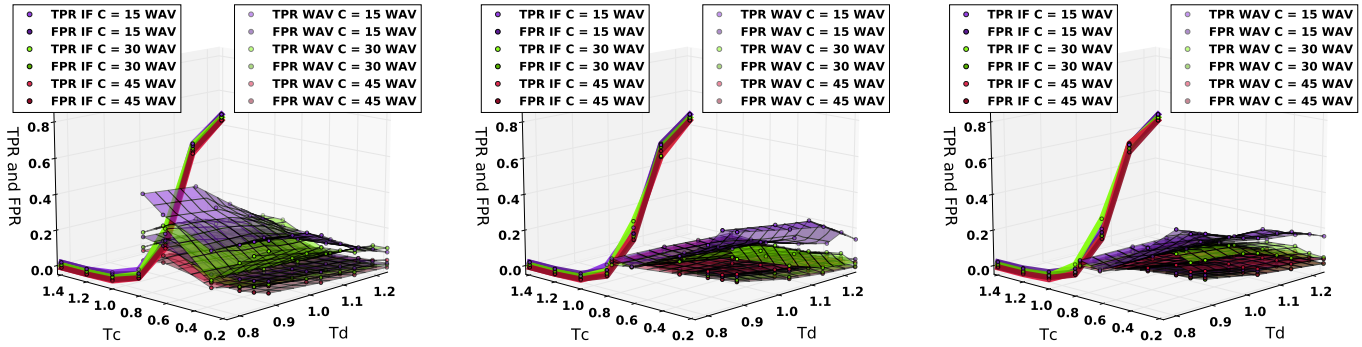
(a) 1 Event (b) 2 Events (c) 3 Events

Fig. 6: $r_i = 7$: Masking Detection ROC curves VS number of events. Performance decrease with more events.



(a) No Event (b) 1 Event (c) 3 Events

Fig. 7: $r_i = 7$: Eliciting Characterisation results. WAV achieves characterisation of $40 - 50\%$ TPR with about 5% FPR.



(a) 1 Event (b) 2 Events (c) 3 Events

Fig. 8: $r_i = 7$: Masking Characterisation results. WAV achieves characterisation of $40 - 50\%$ TPR with about 5% FPR.