## Reviewer 1

The authors propose a classifier that can differentiate the states from a time series of a side channel, such as the electricity consumption or temperature. Action: Make it clear that the side channel is power consumption, not temperature.

The proposed method is a variant of prototype-based classification and builds on a sophisticated matching strategy to work with very little training data.

The authors evaluate their approach on different datasets, where it outperforms other classification methods.

- Investigating side channels for intrusion detection is an interesting and relevant concept. Action: Thank you.
- The evaluation focuses on different datasets and demonstrates the strengths of the proposed matching strategy.
- The novelty is not clear. A large body of work on identifying malicious activities using physical measurements exists. These measurements are not called side channels, yet they are similar in scope. Action: Improve the related work section to highlight the gap that DSD proposes to fill.
- The evaluation is not convincing. No security task is considered, such as intrusion or malware detection. As a result, the actual utility of the approach in practice remains unclear. This is a pity as the presented results look promising. Action: Design a better demonstration to proove the usefullness of detection high level activities from low level data.
- Finally, there are a different issues in the presentation that make the paper hard to follow. The proposed method could be described more accurately and better embed in the body of existing work. Action: Write better I guess…?

This is an interesting paper. Action: Thanks, could have said that with a positiv review but thanks… I like the idea of spotting malicious activities by investigating physical measurements, as it deviates from the classic concepts of intrusion and malware detection. However, I am unsure what to make with this paper, as it suffers from different shortcomings. In the current form, I think that it is not ready for publication at a security conference.

I like the idea of using a side channel to monitor the current state of an electronic device. Yet, I am not really convinced that the paper is adequately positioned and contains sufficient novelty for a top conference. Moreover, there are a few statements in the submission that I need to be clarified with respect to this idea.

- First, the authors argue that side channels are a robust data source that cannot be forged (easily). Is this really the case? Action: Good point. I am missing some theoretical or empirical evidence for this claim. I am skeptical because several side channels can also be used for stealth communication. In this case, the adversary arbitrarily manipulates the channel, for example, a device's electricity consumption, temperature, or electromagnetic emanation. The same techniques could be applied to cloak an attack or at least create false alerts. Action: Find a paper to show that power consulmption is difficult to forge. Maybe talk to carlos about it. Read Carlos papers to see what they claim.

I would appreciate it if the authors could tone down claims on the robustness of side channels unless they can provide clear evidence Action: I'd appreciate if you tone up your review! . The mere fact that manipulating the side channels appears difficult is not sufficient, in my view. Moreover, it would be great if the authors could comment on the relation of side-channel communication concerning robustness.

- Second, I find the discussion of related work very narrow. The authors mainly focus on previous work investigating side channels. This makes sense from the perspective of an attacker but not from the view of a defender. A defender can usually employ a sensor to measure data, similar to a side channel. For example, temperature and electricity consumption are frequently available as hardware sensors in many devices. Action: NO! The whole point is to do that independently from the machine! Read the paper! (note: make that clearer) Consequently, the proposed method is closely related to a large body of work that uses physical measurements for spotting attacks, such as works on intrusion detection in IoT networks, embedded devices, and SCADA networks. Action: NO NO NO we cant use data sent by the device, can't trust it!!

I highly recommend extending the discussion of related work and providing a broader discussion of physics-based monitoring and intrusion detection. Action: yes, do that. In my understanding, the approach is not fundamentally different from other work using available physical sensors Action: make crystal clear that embedded sensors cannot be trusted. Yes they provide more insight than global consumption but at the cost of being tampered with, bypassed, forged, etc. Can't use them! , and thus the authors should clarify its role and better illustrate the technical and operational differences.

The presented evaluation is not really convincing, as it does not investigate a security task, such as intrusion or malware detection. Action: Find a crypto mining malware and detect the shit out of it. Cite a malware that requires a restart to start operation (detect the restart, detect the malware). Cite a malware that prevent windows update. Add a whole related work section about malware behavior. As a result, the actual utility of the proposed method in practice is not evaluated.

- Figure 3 nicely shows that MAD outperforms the other approaches with respect to accuracy. However, in a multi-task setting, the accuracy alone can be misleading, especially if a security task with imbalanced classes is considered. On the one hand, we cannot see the difference between macro-average and micro-average accuracy. That is, what is the performance for each state on average? On the other hand, I am missing performance measures like precision/recall that help to put the prediction of MAD into the context of the classes. I assume that some states dominate the datasets, and therefore both suggestions would enable further insights. Action: add more performances metrics of justify why they are not needed. Computing micro andmacro accuracy is a good idea.
- The conducted experiments have no direct relation to security, particularly attack and intrusion detection. While the proposed method outperforms the baselines, it is unclear whether the reported performance is sufficient for a practical application. For example, an error of 2% in predicting the state of a device might already induce a prohibitive number of false alerts in practice. I would highly recommend expanding the evaluation and adding experiments with security context. Action: talk about false positive, L3 said that they don't mind false positive
- Furthermore, I am wondering about the complexity of the different states. For several datasets, there are only three states (on/off/boot), where the first two should be trivial to discriminate based on the electricity consumption. It would help if the authors could describe their data in more detail and argue why the learning tasks are sufficiently complex. Action: experiment with the detection of more complex states. also argue that with 3 simple states the detection capabilities are already enormous. Provide a chart with some attacks and rules to detect them with simple states.

Finally, I have a couple of questions about the approach. Maybe I overlooked some details in the presentation, but some parts remain unclear to me.

- In Problem Statement 1, the authors define that a sample $t[i]$ maps to a pattern $P_j$. However, they do not specify where $t[i]$ matches within $P_j$. This can lead to ambiguities and makes the map undefined. For example, suppose we have two patterns that overlap $P_a = 123$ and $P_b = 234$. Then, for a time series $t = 1234$ and $t[i] = 2$, both patterns perfectly match. Action: add a sentence to make very clear that the widnow moves and matches all occurences.

- The matching becomes additionally ambiguous due to the used distance metric, which aims to find the best match but does not account for the order of the elements. That is, the patterns $P_a = 321$ and $P_b = 123$ both perfectly match on a time series $t = 222$, as they have the same Euclidean distance. Action: don't understand that.

Maybe this is just a minor issue, but it would be great if the authors could either clarify my misunderstanding or alternatively correct the definitions underlying their approach.

- The run-time complexity of the proposed method looks quite large to me. The authors argue that it is not considerably different from 1 nearest-neighbor classification, which is sometimes called prototype-based classification. However, both techniques have a run-time complexity of $O(P)$ unless some efficient data structures, such as ball or cover trees are used. In this case, the run-time might be reduced to $O(C \log(P)$ for some constant $C$. Action: thats not how the O notation works, take an algorithm analysis class. Still check time efficiency claim.

I would recommend providing more detail on the run-time complexity, including the search for nearest patterns. Moreover, it might help to present run-time numbers in the empirical evaluation, indicating that a good performance can be reached. Action: provide the time it takes for detection and the machine specs.

## Reviewer 2

This paper presents a classification method for time-series data to determine machine activity. The proposed technique was evaluated with several datasets and compared with existing simple classifiers.

Side-channel analysis is valuable in detecting attacks.

- The motivation behind the research is not well established.
- The paper lacks challenges related to physics-based security. Action: same a before, find some better experimental setup
- The proposed technique does not offer significant novelty. Action: your review does not offer significant novelty

The paper appears to be in an initial and incomplete state. Action: your review is in an initial AND incomplete state There seems to be a lack of clear connections between the paper's subject matter, which includes side-channel analysis and physics-based security. It seems that the paper focuses on studying a simple classifier for given time-series data without presenting any noteworthy challenges specific to the problem.

Furthermore, it is unclear how the proposed technique contributes in comparison to existing methods Action: there is no other existing method . The paper should provide unique challenges and valuable insights to effectively address the problems.

Minor: It is unclear from Figure 3 where exactly the proposed technique is depicted, and the meaning of "DSD" is not clarified. Action: update figure 3 to remove DSD.

## Reviewer 3

This paper proposes a mechanism to detect machine activity based on side channel information such as power consumption. Corresponding side channel information can be collected with low intrusiveness (requiring no changes to deployed systems) and is difficult to forge. Thus, leveraging it to enforce security policies is a promising strategy. This paper specifically proposes an approach that requires only one training sample for each type of device activity, which makes the collection of data required to build the underlying model significantly easier.

The approach does not require any changes to monitored machines and can be realized completely separated from the observed network, which makes deployability significantly easier than other detection approaches and at the same time makes it difficult for adversaries to circumvent detection, e.g., by forging measurements. The approach requires only one training sample for each class, significantly easing the creation of suitable training data.

- The evaluation mainly focuses on accuracy and similarity of state predictions (as measured by a condensed version of Levenshtein distance), while the related field of anomaly detection primarily considers metrics such as precision which also incorporate false positives. Action: make extra clear that precision alone is not enougth when the output is state detection and exact precision is not important.
- The example used in the discussion and also the selection of activities in the evaluation are rather simple, making it difficult to imagine whether the approach can also be used for more complicated scenarios. Action: good point, again find better experiment scenarios.
- The approach seems to require a lot of manual tuning by human expert, raising the question whether it is really easily deployable. Action: yes, further work should solve this problem, maybe mention that.
- The paper lacks a lot of interesting details and would have benefited tremendously be providing more context and background information. Surprisingly, this would have been easily possible as the paper is significantly below the maximum page length.

This paper proposes an interesting idea by leveraging side channel information collected outside the actually monitored system to detect system activities which can then be used to test compliance with higher-level security policies. Initially, I was really excited as my assumption was that the approach strives to detect non-trivial activities that can not be captured by very basic heuristics. Much to my disappointment, it appears, however, that the approach essentially can only detect extremely high-level information such as “computer is powered on” (based on observed power consumptions). If this is the intended use of the proposed system, I wonder whether simply defining heuristics, e.g., based on thresholds for power consumption, wouldn’t produce similar or even better results? I might have misunderstood the core contribution of this paper, but in the end I had the impression that the paper performs a lot of complicated steps to achieve results that seem to not require such a complicated approach. Most importantly, I do not believe that the paper actually shows that “side-channel based state detection enables a more robust security policy enforcement”. To prove this claim, I would have expected a significantly more involved evaluation alongside a set of defined security policies which can be enforced “better” using the approach presented in this paper compared to related work and especially simple threshold-based heuristics.

Regarding the evaluation, I could not follow the rational behind the selection of performance metrics. In the field of anomaly detection, besides attempting to uncover as many as possible anomalies as possible, one of the main goals is to reduce the number of false alarms. This is especially relevant also in the context of this paper, as in reality the number of benign events will be orders of magnitude larger than the number of malicious events. Consequently, even the smallest tendency to produce false alarms will inevitable lead to a situation where virtually all alarms are false alarms, resulting in alarm fatigue of human operators. Action: better explain the performances metrics Consequently, it is extremely important to use performance metrics such as precision which also consider false predictions. Action: make extrac clear that precision is not relevant here because the output is not individual labels of sample but SEGMENTS of label. Where the segment start and stop is not so important compared with its very existence. I a sense, the paper tries to address this issue by calculating Levenshtein distances on a condensed list of state predictions, but it remains unclear how this custom metric can be used to ensure that the approach does not produce any / too many false alarms. Especially to also cover temporal aspects, it might be promising to also consider enhanced time-aware (eTa) metrics [A], especially eTaP (precision). Action: look into that. The question of the correct choice of metrics is especially pressing, since the approach presented in this paper only clearly outperforms other (generic machine learning) approaches for the tailored version of the Levenshtein distance.

I believe at least some of my concerns could have been avoided by simply providing more information and context in the paper (given that it is by far below the page limit). For example, I would have liked to read a thorough introduction into the research field that is clearly separated from the discussion of related work. In turn, I believe the discussion of related work could have been significantly extended, e.g., by covering the vast research area of anomaly detection in cyber-physical systems (often evaluated on the SWaT and WADI datasets). Most importantly, we do not learn about approaches relying on “traditional sliding window algorithm” or “traditional 1-Nearest Neighbor (1-NN)”, despite those appearing to be the closest related work. Action: add that in the related work section, talk about sliding windows methods. Overall, I missed a lot of explanations and high-level intuition throughout the paper, which made it quite difficult to figure out which precise problem the paper attempts to tackle and how exactly the proposed approach works. In a sense, I had the feeling of reading a paper that is half finished, which is a pity given the extremely promising idea underlying this paper. Action: in the intro, provide the intuition for the DSD

Minor comments: Figure 1 seems to not be referenced / explained in the text. Action: unacceptable, fix that It remains unclear why time-efficiency and termination are exactly those two properties that warrant a dedicated analysis in Section 4.2. Figure 3 refers to “DSD”, when likely “MAD” is meant. (Likely) “0” missing in “215” in the description of the REFIT dataset.

References: [A] Hwang, W.S., Yun, J.H., et al.: “Do You Know Existing Accuracy Metrics Overrate Time-Series Anomaly Detections?”. In: ACM SAC (2022)