WILLIAM & MARY
CHARTERED 1693

CSCI 780: IoT Security

Prof. Adwait Nadkarni

Lecture 2

Security Research Methods: Reading Papers

Why do we read papers?

Understanding a paper

- **Q1]** What is the *central idea*?
  - Abstract
  - Introduction
  - Conclusion

  These are the best areas to find an overview of the contribution

- **Q2]** Where does the work *fit* in the area?
  - *Ideally*, the *related work* section should describe this well.
  - Papers that do not do this or do a superficial job are almost sure to be bad ones
  - An informed reader should be able to read the related work and understand the basic approaches in the area, and how they differ from the present work

Understanding a paper

- **Q3]** What *claims* do the authors make? (examine the abstract, intro, conclusion for high-level claims, the "design/analysis" section for more precise claims)
- **Q4]** What *scientific devices* are the authors using to communicate their point or evaluate their claims?
  - i.e., the *methodology*
  - *Theoretical papers* validate a model using a mathematical argument (i.e., a proof)
  - *Experimental* papers use a test apparatus to evaluate claims (e.g., performance of a detection system under simulated workload)
  - *Empirical* claims are evaluated via measurement

Understanding a paper

- **Q5]** What did they find?
  - Results - statement of new scientific discovery.
    - Typically some abbreviated form of the results will be present in the abstract, introduction, and/or conclusions
    - *Note*: just because a result was accepted into a conference or journal does necessarily not mean that it is true. Always be circumspect.
- **Q6]** What should you remember from this paper, i.e., what is the *takeaway*?
  - i.e., what general lesson or fact should you take away?
  - Really good papers have takeaways that are more general than the paper topic.

The best papers are the ones that teach you something

Reading Tips

- Everyone has a different way of reading a paper
- Here are some tips for effective reading (and recollection):

  1. Always have a copy to mark-up (Digitally, use Zotero/ Mendeley)

  2. After reading, **write a short summary of the paper**, and ideally, keep it with the paper. The summary should contain the following points:

    - area, problem, solution, methodology, results, takeaway, and key questions/ideas emerging from the paper.

The security publishing model

Derived from slides by William Enck

Where to Publish

- Traditional Venues:
  - Journals
  - Conferences
  - Workshops
- Tech Reports (*i.e., self-publish*)
- Books (less frequent, more work)
  - Book chapters (more frequent than books)

Archival

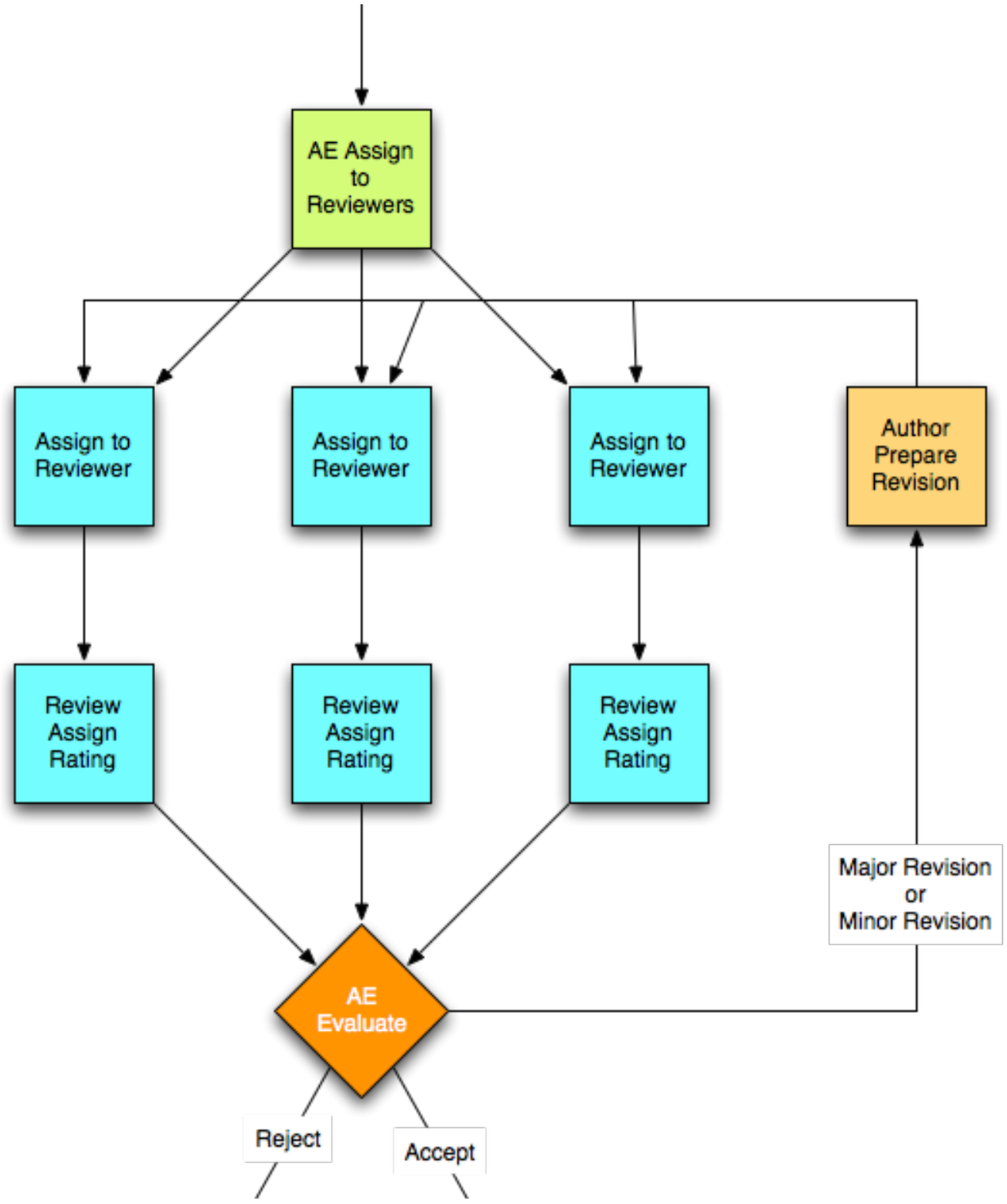
Preliminary

Publication Tiers

- Not all venues are the same
- **Tier 1 (i.e., *top-tier*):** IEEE S&P/Oakland, USENIX Security, ACM CCS, ISOC NDSS, TOPS (*journal*), JCS (*journal*)
- **Tier 2:** ACSAC, ACNS, ESORICS, CSF, AsiaCCS, TOIT (*journal*)
- **Tier 2.5/3:** CODASPY, SecureComm, WiSec, RAID,
- **Tier 4:** HICS
  - SCIgen (WMSCI 2005)
  - http://pdos.csail.mit.edu/scigen/

Journal Publication

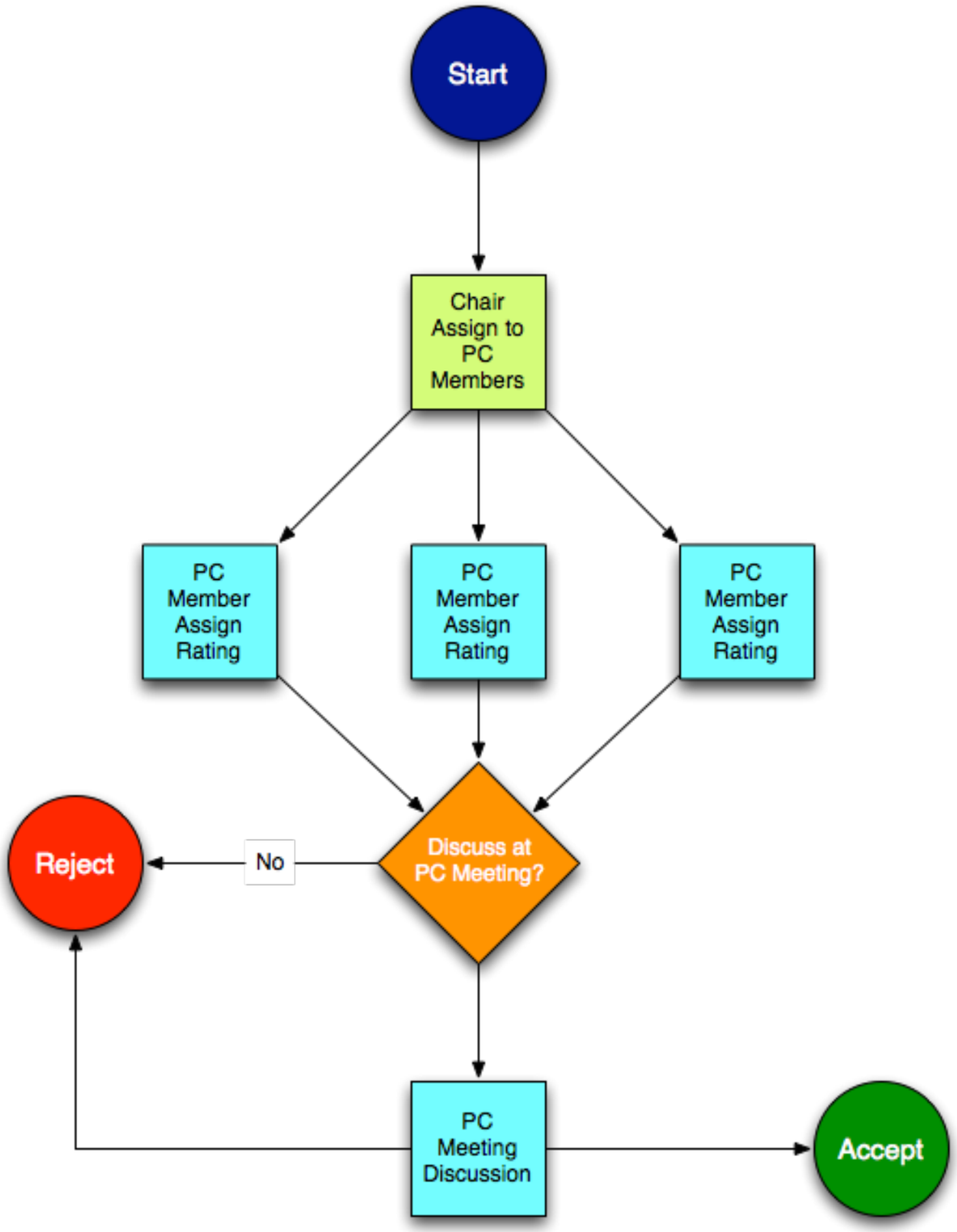
- The editor-in-chief (EIC) receives the papers as they are submitted.
- The papers are assigned to associate editors for handling.
- Anonymous reviewers rate the paper:
  - Accept without changes
  - Minor revision
  - Major revision
  - Reject



- *CCS, and NDSS are also using such ratings now,* ***Stopped in S&P 2024, USENIX 2026***

Conference Publication

- The PC Chair is the person who marshals the reviewing and decisions of a conference.
- This is different than the general chair.
- PC members review, rate and discuss, the paper, then vote on which ones are accepted.
- The acceptance rate is the ratio of accepted to submitted papers.
- Conferences may also use *area chairs, review task forces (RTFs) or shadow PCs* to ensure the quality of the reviewing process.



Evaluating a Paper

- A paper is evaluated on:
  - Novelty
  - Impact
  - Correctness
  - Presentation
  - Relevance
- “hotness” may also factor into the reviews

*Which of these is an **objective** metric?*

Peer-Reviewing Papers

Why do we *review* papers?

Reviewing is *service*

Two critical functions

Assessment/QC

Helping fellow researchers

Reviews: The good, the bad, and the ugly

- Review others' work like you want yours to be reviewed.
- Appreciate good research, don't nitpick, and be constructive.

Detailed comments for authors

I enjoyed reading this paper, and I think it contains some interesting ideas. Although I do not support acceptance at this time for the reasons specified below, I could imagine a nice publication resulting from this line of work (or perhaps even two publications if you decide to separate out some of your findings on generating sequences).

Paper Focus and Sequence Generation Evaluation

When I was a third of the way through reading this paper, I made a note to myself that I remained uncertain what the paper would actually do. Ultimately, more than half of the paper is devoted to presenting a method for generating likely sequences of events for home automation systems. This feels like far too much space devoted to explaining relatively straightforward concepts (I do not say "straightforward" as a criticism of your approach; my point is that a much more concise presentation seems feasible). The focus of the paper seems to be on sequence generation. To the extent that generating likely sequences has a relationship to security, those considerations feel like an afterthought in the paper.

Furthermore, the authors ultimately skip over details regarding the quality of the resulting sequences. Plenty of promising systems for text generation still occasionally produce humorously bad results (for example, see email auto-response suggestions). I worry about the imp
prone system here. I would like to be convinced that your system cons
sequences. One approach would be to have users manually validate t
reasonable.

I would recommend that you dramatically compress your explanation
method, add more testing to show that it generates reasonable seque
to discussing and demonstrating its importance for security.

Also note that Section 6 takes an abrupt turn. Earlier sections left me
evaluation focus would be on normal users, but this section focuses o
vendors.

A (USENIX'20)

There has been a great degree of interest in security and safety policies for smart homes in recent years. While a number of papers cited by this work have sought to create and enforce security properties for such situations, the evaluation of these properties has always been somewhat incomplete because few large, real-world data sets from actual homes are available. This paper aims to fix that, which is a laudable goal. The core problem is that what this paper accomplishes is not sufficiently real, and also perhaps not sufficiently novel, to achieve that goal.

The paper claims that analysis of Samsung SmartThings apps is the state of the art for home automation data sets. However, this is not quite true. Some recent home automation papers with similar goals, like [33], used large-scale data collected from IFTTT:

- Mi et al. An empirical characterization of IFTTT: ecosystem, usage, and performance. From IMC '17 https://dl.acm.org/citation.cfm?id=3131369
- Ur et al. Trigger-Action Programming in the Wild: An Analysis of 200,000 IFTTT Recipes. From CHI '16 https://dl.acm.org/citation.cfm?id=2858556

This paper collects IFTTT-like trigger action programs in a small user survey. This paper's technique has the advantage that a collection of apps comes from a single user, whereas those large IFTTT data sets only contain the name of the person who created the app, rather than everyone who is using it. As such, this paper could make a great contribution. Unfortunately, the apps users in this study make don't control anything in the real world. Users don't have the chance to iteratively add to or improve their rules. In addition, the traces are simulated based on a language model based on these rules. Thus, it seems like this artificial approach does not have enough ecological validity or external validity to approximate real traces. Furthermore, based on the
ntered manually for these rules, rather than
le events. In other words, the interface in the
used systems like IFTTT.

at Washington State from a decade ago also
u.edu/datasets/

security and safety violations are very rare.
guage, why not use a formal model? Much of

B (CCS'20)

Comments for author

This paper is difficult to follow. The motivation is not quite clear. According to Section 2, Helion could avoid the put of "tremendous amount of time and effort". However, on the other hand, a machine learning-based approach requires "the existence of a set of natural scenarios" as training data, which also needs many efforts. Such motivation is not convincing enough.

Also, Helion does not consider the possibility of personalized settings. It implies that every user has the same need. Further, the output of Helion relies on the quality and quantity of input data.

The evaluations are not convincing. In Section 8.1, no quantitative analysis and horizontal comparison were conducted. The assessment is only based on the judgments of the authors. Such evaluation cannot demonstrate the advancement of Helion, compared with the previous solutions. In Section 8.2, similarly, why the previous solution cannot achieve the same result?

Finally, it seems this paper is not a security paper. The security-related contents/contributions are just a small part of this paper, and these contents could be removed without affecting the logic. Maybe this paper is more suitable for ubiquitous computing or human-computer interaction related conferences, like UbiComp and CHI.

Some small issues:

(1) What is the meaning of "Helion"? It is a bit confusing.

C (also CCS'20)

Common Problems/Myths

1. This is a published, award-winning paper, so there must be *no weaknesses!*

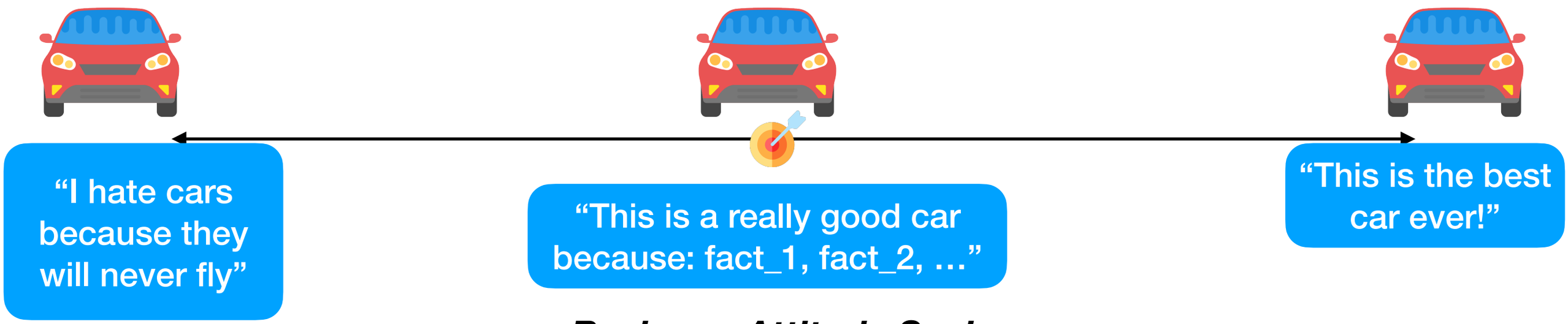
- This is a trap.
- Some of the best papers have weaknesses/gaps/limitations
- Weaknesses are *nothing personal*
- Rather, identifying weaknesses often leads to future research

Think critically, and be skeptical

Common Problems/Myths

2. This is an unpublished paper, so it must not have any strengths.

- This is a trap, as well. Don’t expect the impossible.
- Several Influential papers have been published at venues that were not “top-tier”, or have existed outside of peer-review entirely!
- Be *positive* when reviewing a paper



Reviewer Attitude Scale

Common Problems/Myths

3. Not *justifying* your position

- Whether your position is positive or negative
  - *Justify it with facts*
- If its an opinion,

  1. Be forthright and clearly state so

  2. Be ready to change it when presented with facts to the contrary (e.g., in the discussion, rebuttal)

- Where to justify? *Generally,* <u>“Detailed Comments for Authors”</u>

Common Problems/Myths

4. No *constructive* comments

- **Recall:** *Reviewing is about helping your fellow researcher*
- Be honest, but constructive
- Which of these would *you* like to receive for your paper?

  a. “The paper does not have property X” OR

  b.“The paper should look in Y direction, which will potentially help it achieve property X”

- Always *back up your comments with facts*, e.g.,

  a. “The paper should compare its performance (or security properties, or …) with related works [1] [2] [3] (*references provided below the review*).” *is much better than*

  b.“The paper does not compare itself with related work adequately”

On weaknesses

- Is it really a weakness *of the paper?*
  - General weakness: “the paper tests with few apps”
  - Weakness in the context of the paper’s goal: “the paper is about evaluating X, which is why it should be tested with more applications”
  - Weakness in the context of the paper’s goal, and claims: “The paper evaluates X, and claims Y about it, which can only be sound if evaluated with a larger set of apps”

Reviews: The good, the bad, and the ugly

- Review others' work like you want yours to be reviewed.
- Appreciate good research, don't nitpick, and be constructive.

Detailed comments for authors

I enjoyed reading this paper, and I think it contains some interesting ideas. Although I do not support acceptance at this time for the reasons specified below, I could imagine a nice publication resulting from this line of work (or perhaps even two publications if you decide to separate out some of your findings on generating sequences).

Paper Focus and Sequence Generation Evaluation

When I was a third of the way through reading this paper, I made a note to myself that I remained uncertain what the paper would actually do. Ultimately, more than half of the paper is devoted to presenting a method for generating likely sequences of events for home automation systems. This feels like far too much space devoted to explaining relatively straightforward concepts (I do not say "straightforward" as a criticism of your approach; my point is that a much more concise presentation seems feasible). The focus of the paper seems to be on sequence generation. To the extent that generating likely sequences has a relationship to security, those considerations feel like an afterthought in the paper.

Furthermore, the authors ultimately skip over details regarding the quality of the resulting sequences. Plenty of promising systems for text generation still occasionally produce humorously bad results (for example, see email auto-response suggestions). I worry about the imp
prone system here. I would like to be convinced that your system cons
sequences. One approach would be to have users manually validate t
reasonable.

I would recommend that you dramatically compress your explanation
method, add more testing to show that it generates reasonable seque
to discussing and demonstrating its importance for security.

Also note that Section 6 takes an abrupt turn. Earlier sections left me
evaluation focus would be on normal users, but this section focuses o
vendors.

A (USENIX'20)

There has been a great degree of interest in security and safety policies for smart homes in recent years. While a number of papers cited by this work have sought to create and enforce security properties for such situations, the evaluation of these properties has always been somewhat incomplete because few large, real-world data sets from actual homes are available. This paper aims to fix that, which is a laudable goal. The core problem is that what this paper accomplishes is not sufficiently real, and also perhaps not sufficiently novel, to achieve that goal.

The paper claims that analysis of Samsung SmartThings apps is the state of the art for home automation data sets. However, this is not quite true. Some recent home automation papers with similar goals, like [33], used large-scale data collected from IFTTT:

- Mi et al. An empirical characterization of IFTTT: ecosystem, usage, and performance. From IMC '17 https://dl.acm.org/citation.cfm?id=3131369
- Ur et al. Trigger-Action Programming in the Wild: An Analysis of 200,000 IFTTT Recipes. From CHI '16 https://dl.acm.org/citation.cfm?id=2858556

This paper collects IFTTT-like trigger action programs in a small user survey. This paper's technique has the advantage that a collection of apps comes from a single user, whereas those large IFTTT data sets only contain the name of the person who created the app, rather than everyone who is using it. As such, this paper could make a great contribution. Unfortunately, the apps users in this study make don't control anything in the real world. Users don't have the chance to iteratively add to or improve their rules. In addition, the traces are simulated based on a language model based on these rules. Thus, it seems like this artificial approach does not have enough ecological validity or external validity to approximate real traces. Furthermore, based on the
ntered manually for these rules, rather than
le events. In other words, the interface in the
used systems like IFTTT.

at Washington State from a decade ago also
u.edu/datasets/

security and safety violations are very rare.
guage, why not use a formal model? Much of

B (CCS'20)

Comments for author

This paper is difficult to follow. The motivation is not quite clear. According to Section 2, Helion could avoid the put of "tremendous amount of time and effort". However, on the other hand, a machine learning-based approach requires "the existence of a set of natural scenarios" as training data, which also needs many efforts. Such motivation is not convincing enough.

Also, Helion does not consider the possibility of personalized settings. It implies that every user has the same need. Further, the output of Helion relies on the quality and quantity of input data.

The evaluations are not convincing. In Section 8.1, no quantitative analysis and horizontal comparison were conducted. The assessment is only based on the judgments of the authors. Such evaluation cannot demonstrate the advancement of Helion, compared with the previous solutions. In Section 8.2, similarly, why the previous solution cannot achieve the same result?

Finally, it seems this paper is not a security paper. The security-related contents/contributions are just a small part of this paper, and these contents could be removed without affecting the logic. Maybe this paper is more suitable for ubiquitous computing or human-computer interaction related conferences, like UbiComp and CHI.

Some small issues:

(1) What is the meaning of "Helion"? It is a bit confusing.

C (also CCS'20)

Good Luck!