The Ephemeral Threat: Attacking Algorithmic Trading Systems powered by Deep Learning

DECISION: Accept

AUTHORS: Advije Rizvani, Giovanni Apruzzese and Laskov Pavel

Summary of Reviews

- Review 1: 0 (3)
- Review 2: 2 (3)

Reviews

Review 1

Total score: 0

Overall evaluation: 0

Reviewer's confidence: 3

The authors study an overlooked aspect of security research, i.e., time series forecasting of financial predictions. They highlight the vulnerability of a DL-powered algorithmic trading system (ATS) to adversarial perturbations.

I enjoyed reading the paper. The work is very interesting and the threat model along with the concept of ephemeral perturbations (EP) is highly intriguing. The authors show that perturbations (such as a couple of changes in the input stock prices) can lead the ATS to make bad decisions leading to losses.

I have some comments regarding the work and the way the things are explained which need to be addressed before the paper is ready to be published.

1) From Related Works. The authors mention the snowball effect and their "results dont change" with it. The concept lacks explanation in the work. Can the authors please elaborate more on the snowball effect, how they used it and what it means in the context of the study.

2) From Related Works. The authors say:

"Despite providing significant contributions to the state of the art, these six papers ([17–22]) have two important limitations from a security viewpoint. Powerful Attacker: four papers [17–21] envision an adversary who possesses extraordinary knowledge (e.g., for PGD attacks) or capabilities—which

may not reflect a realistic scenario [2]. Model only: five papers [18–22] consider the effects of the attack on the model in isolation—without considering the impact on the overarching system [2]. Finally, only two papers [18, 21] release their source-code”

-> If [17-22] are 6 papers, then how is [17-21] 4 papers? Also, [18-21] should be 4 papers not 2? I am a little confused here.

3) The authors do not discuss Figure 3 in the manuscript. They go over the details of the ATS they developed in Section 4 and associated details, which I appreciate. However, there are graphs in Figure 3 that need to be discussed.

4) In terms of motivation for such an attack. The authors say the following:

“Our attacker has one goal: induce the targeted organization to gain less money”.

I am curious what incentive do the authors think an attacker has to develop systems that can send these perturbations and then fiddle with ATS systems. Are these competitors trying to take down other systems? Has there ever been a case where such attacks have been found in the wild? It would be good to see some discussion along the lines to highlight the relevance and/or broader implications of the issue.

5) Might be out of scope, but it would also be interesting to see some discussion about the "Potential Negative Outcomes" of this work, such as an adversary learning these strategies and using it against systems. And what ATS system owners can do to protect themselves in a real-world scenario?

Review 2

Total score: 2

Overall evaluation: 2

Reviewer’s confidence: 3

The idea of an ephemeral threat is quite interesting. The early results look good.

Questions to consider for further development:

1) what is the attacker's motivation/gain? What's in it for them to manipulate the system?

2) the attacker needs to have access at the brokerage to make the injection practical. How would that be accomplished?

3) the assumption is that there is a single data source (brokerage). It seems that multi-sourcing this information would negate this attack completely.

4) are 38 stocks representative of a portfolio? It appears that a larger portfolio will have vanishing effect of the attack