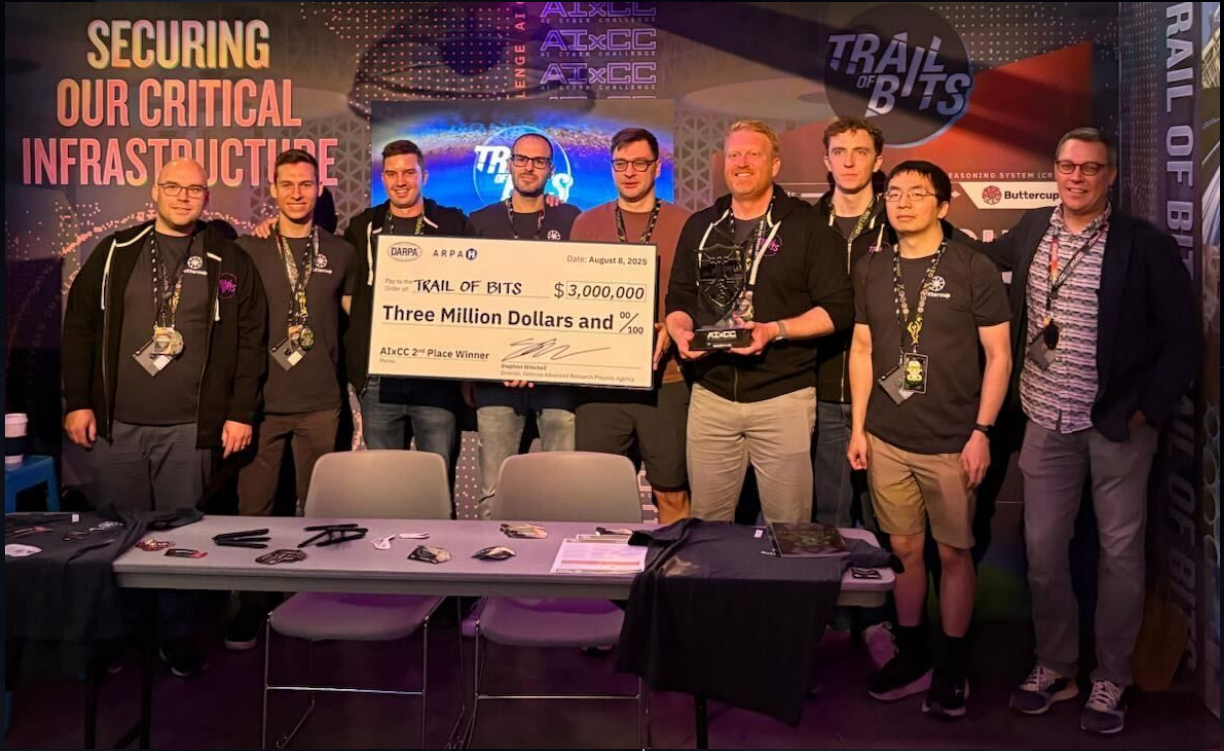
TRAIL OF BITS

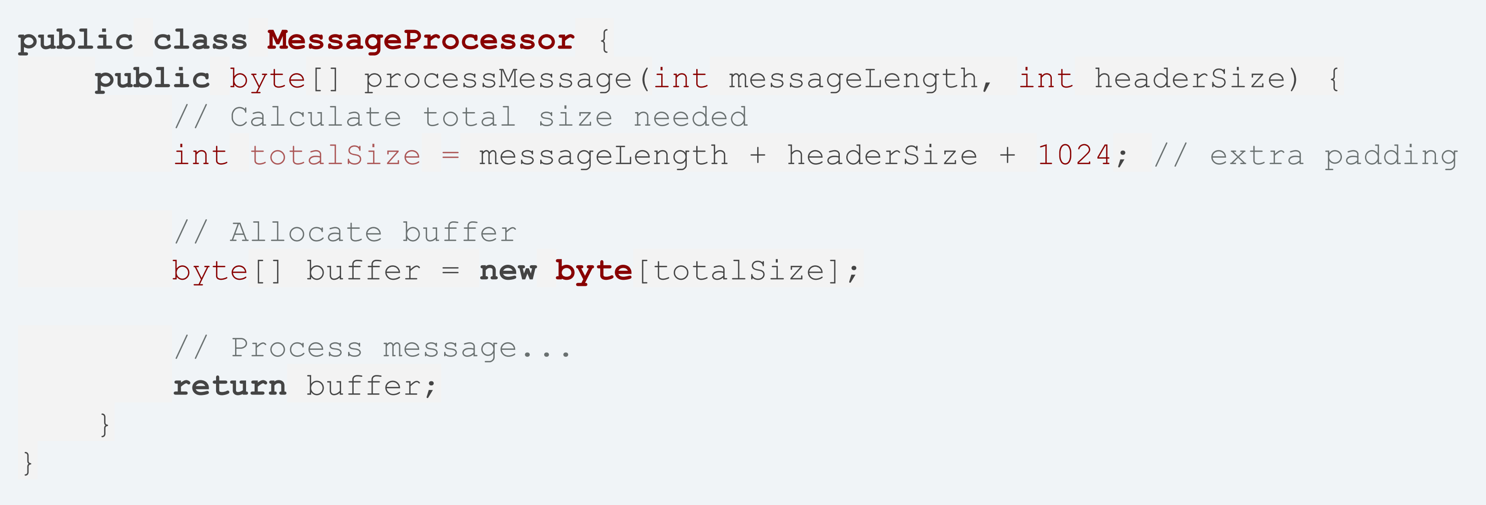


Buttercup: The Future of Trail of Bits' Solution to the DARPA's AI Cyber Challenge

:(
Your PC ran into a problem and needs to restart. We're just collecting some error info, and then we'll restart for you.
100% complete
Clear Channel
2
4

2024 CrowdStrike IT outage

What does a bug look like?

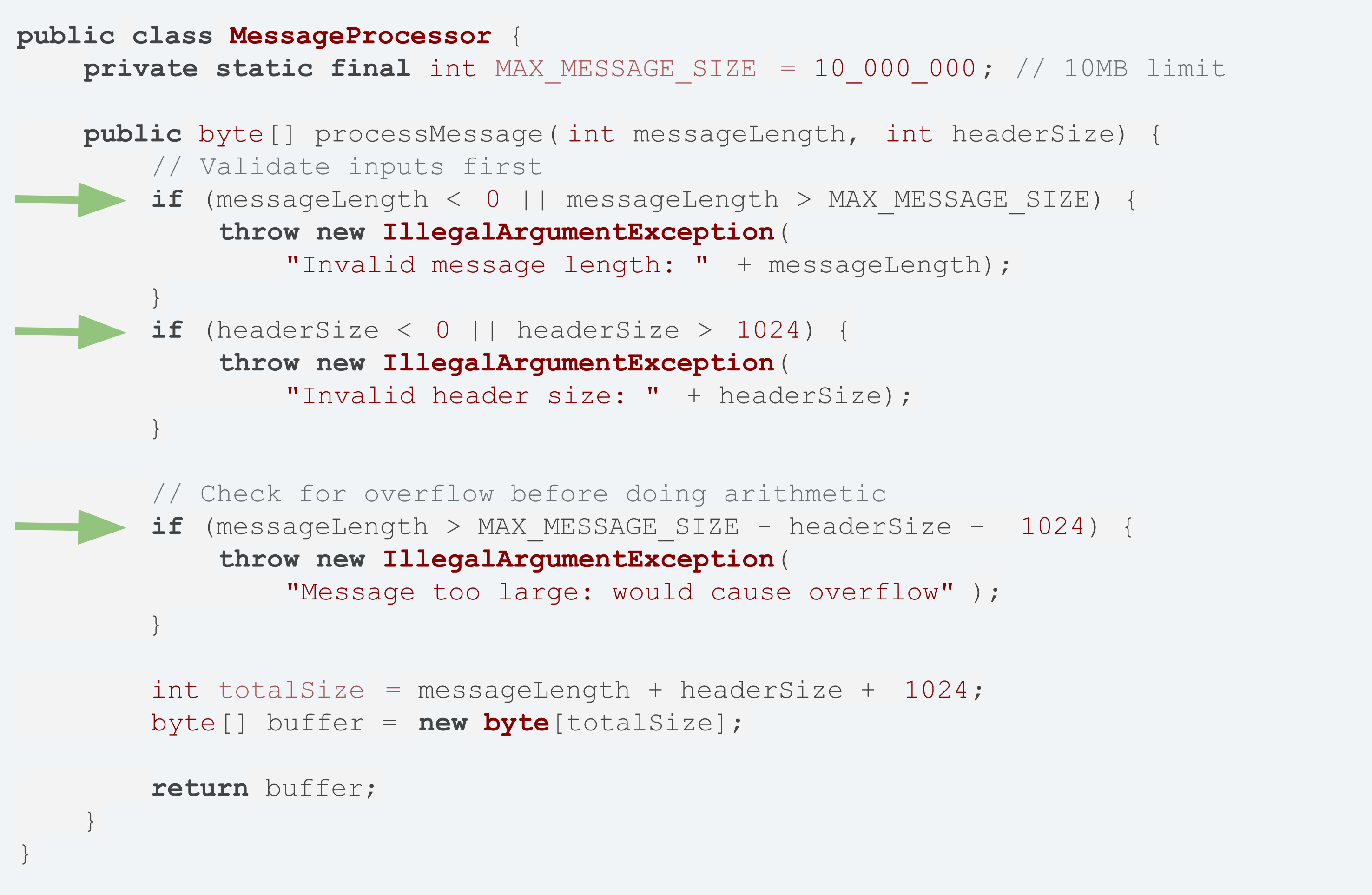


```
public class MessageProcessor {
    public byte[] processMessage(int messageLength, int headerSize) {
        // Calculate total size needed
        int totalSize = messageLength + headerSize + 1024; // extra padding

        // Allocate buffer
        byte[] buffer = new byte[totalSize];

        // Process message...
        return buffer;
    }
}
```

What does a patch look like?



```java
public class MessageProcessor {
    private static final int MAX_MESSAGE_SIZE = 10_000_000; // 10MB limit

    public byte[] processMessage(int messageLength, int headerSize) {
        // Validate inputs first
        if (messageLength < 0 || messageLength > MAX_MESSAGE_SIZE) {
            throw new IllegalArgumentException(
                "Invalid message length: " + messageLength);
        }
        if (headerSize < 0 || headerSize > 1024) {
            throw new IllegalArgumentException(
                "Invalid header size: " + headerSize);
        }

        // Check for overflow before doing arithmetic
        if (messageLength > MAX_MESSAGE_SIZE - headerSize - 1024) {
            throw new IllegalArgumentException(
                "Message too large: would cause overflow");
        }

        int totalSize = messageLength + headerSize + 1024;
        byte[] buffer = new byte[totalSize];

        return buffer;
    }
}
```

Key Scoring Components

Points Awarded For:

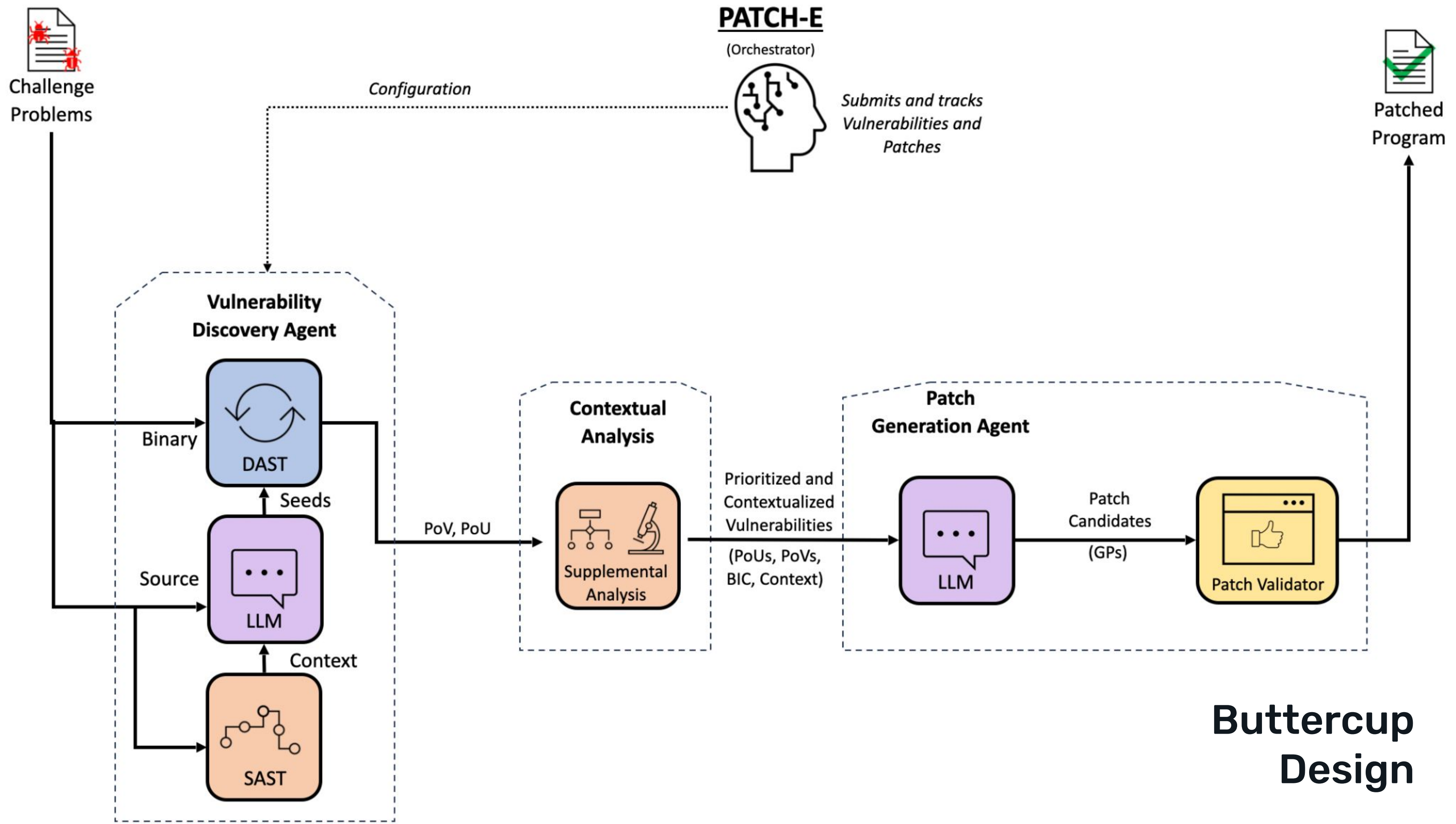
- **Patches** - Worth the most points
- **Proof of Vulnerabilities** - Worth moderate points
- **Bundled submissions** - Extra points when POV + patch submitted
- **Static analysis reports** - Minor points (mainly useful when bundled)

Scoring Modifiers:

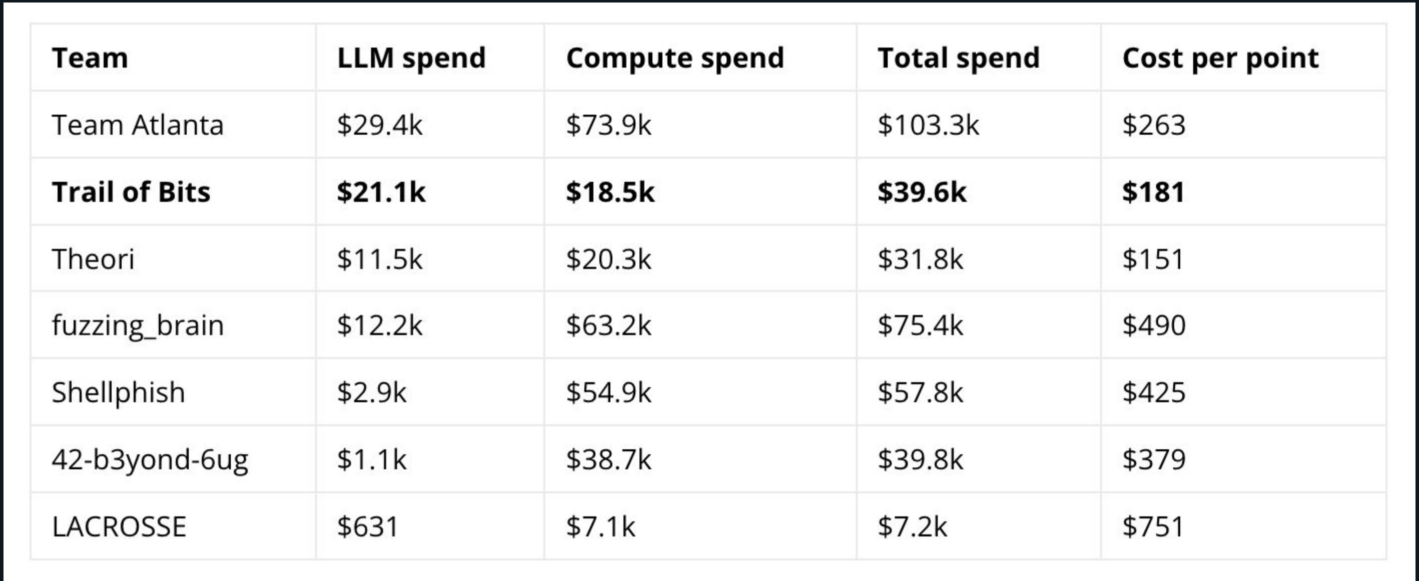
- **Speed bonus** - Faster submissions earn more points
- **Accuracy multiplier** - Incorrect/duplicate submissions reduce your score multiplier
- **Baseline comparison** - Must outperform state-of-the-art baseline systems

Penalties:

- Duplicate POVs or patches reduce accuracy multiplier
- Incorrect patches (that don't fix the bug or break functionality) harm accuracy multiplier
- Submitting patches that get "clobbered" by later POVs loses points



Buttercup Design



Team	LLM spend	Compute spend	Total spend	Cost per point
Team Atlanta	$29.4k	$73.9k	$103.3k	$263
Trail of Bits	**$21.1k**	**$18.5k**	**$39.6k**	**$181**
Theori	$11.5k	$20.3k	$31.8k	$151
fuzzing_brain	$12.2k	$63.2k	$75.4k	$490
Shellphish	$2.9k	$54.9k	$57.8k	$425
42-b3yond-6ug	$1.1k	$38.7k	$39.8k	$379
LACROSSE	$631	$7.1k	$7.2k	$751

Results!

2nd in LLM Spend

6th in Compute Spend

2nd in $ per Point

The Future of AI in Cybersecurity

- Determining exploitable vulns
- Binary reverse engineering
- Formal methods reasoning
- No more malware C&Cs
- 🤫 No more SBIR Phase I’s

trent@trailofbits.com
www.trailofbits.com/buttercup

Our work, blog.trailofbits.com
Our code, github.com/trailofbits
Our socials, @trailofbits on X