

Natural Language Processing Methods for Software-Based Leak Detection, Prevention, and Mitigation

Jacob Fuehne

NLP Researcher
jfuehne2@illinois.edu

Vivek Nair

Authentication Researcher
vcnair@solidsecurity.co

Project Tailspin

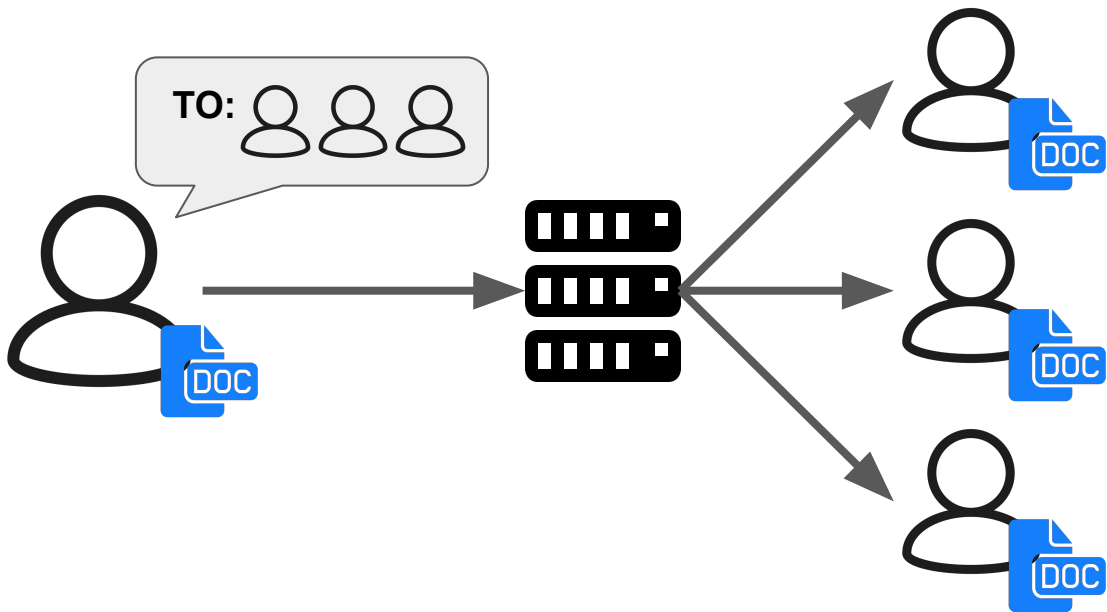


Project tailspin was created to focus on leak mitigation for text-based documents.

Before Tailspin

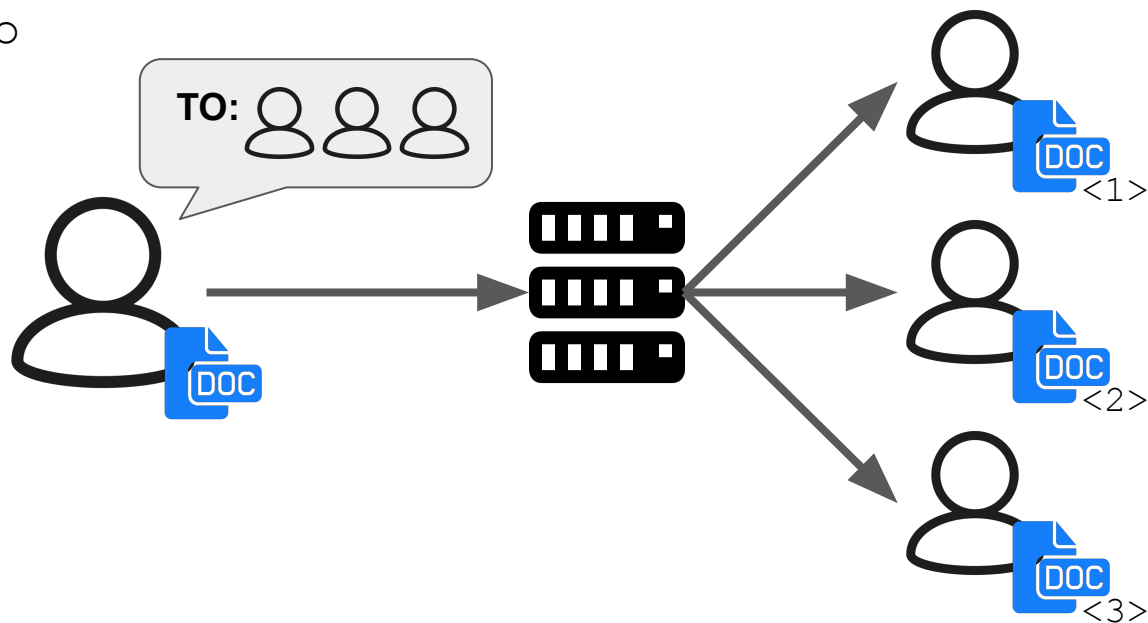
Before { | } PROJECT
TAILSPIN

Original documents are distributed directly to intended recipients.



Metadata Tagging

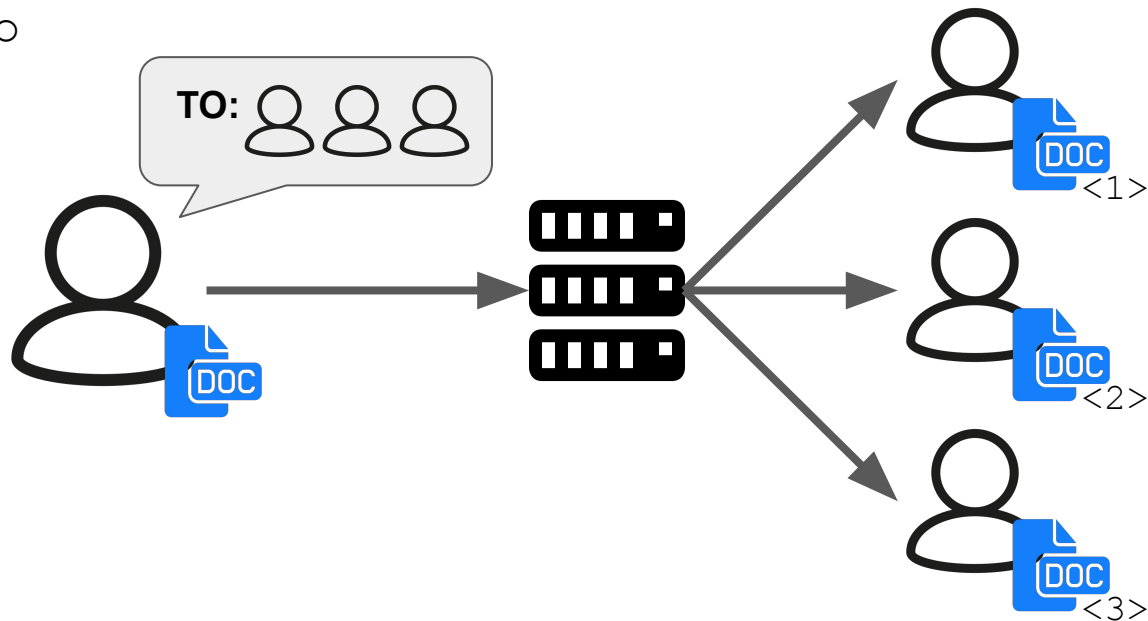
Original documents are tagged with unique metadata before being distributed directly to intended recipients.



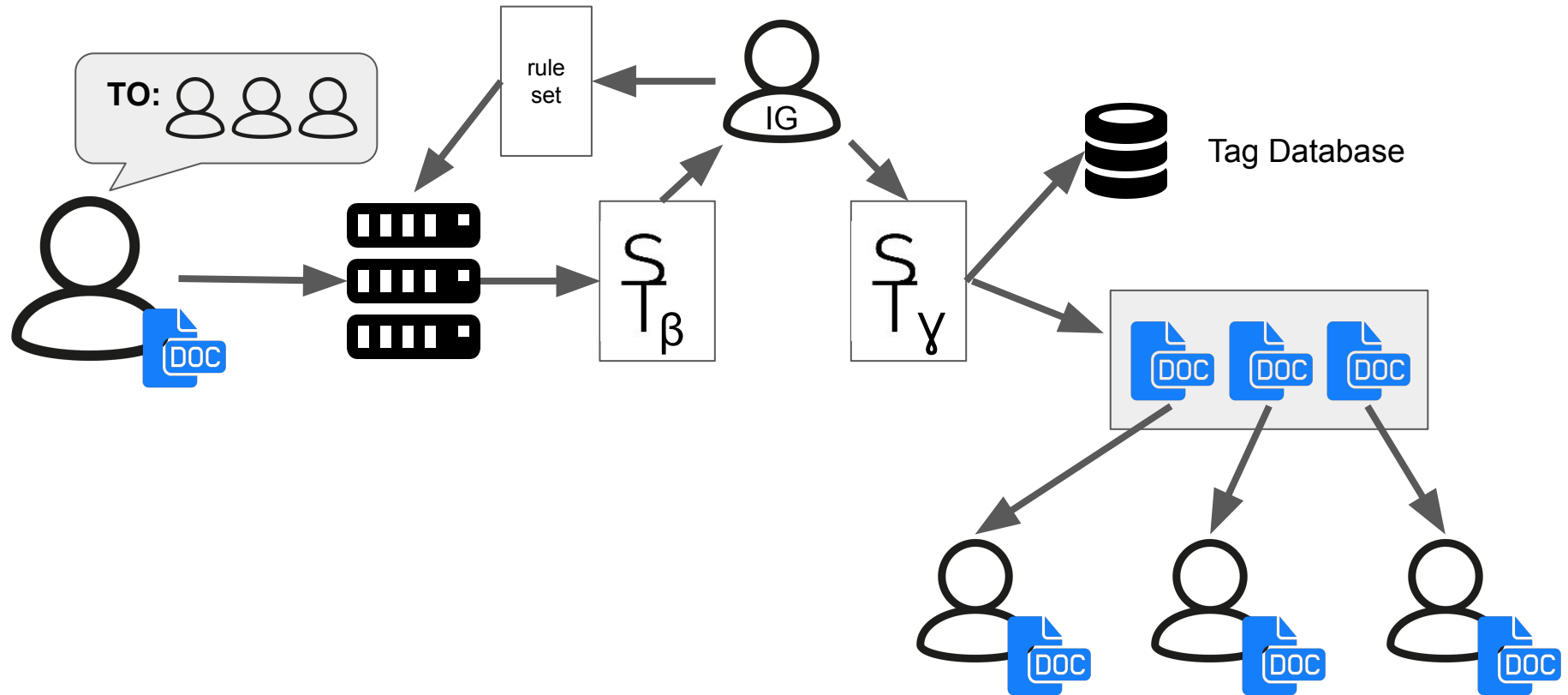
Metadata Tagging

Original documents are tagged with unique metadata before being distributed directly to intended recipients.

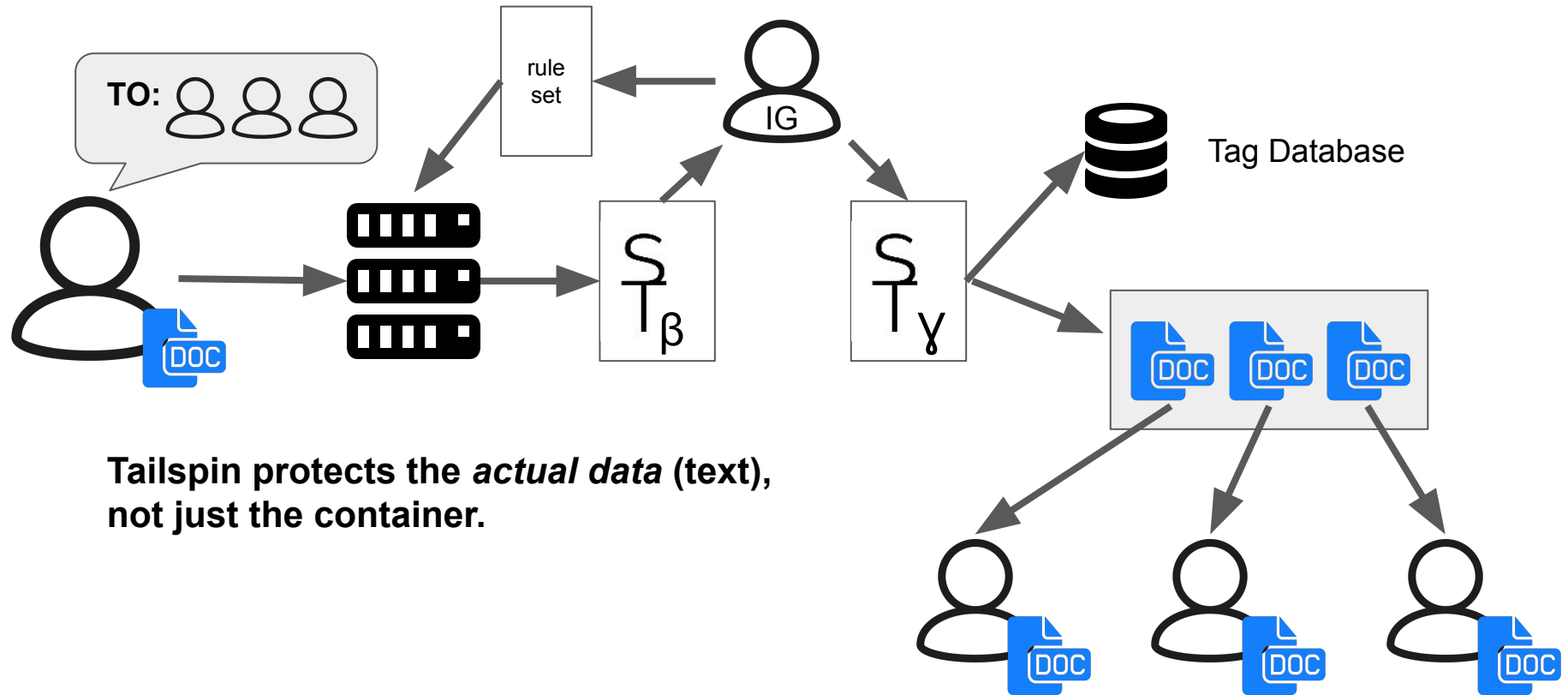
This method protects the *container*, not the *"real data"* content.



With Tailspin



With Tailspin





Objective 1: use NLP techniques to make text uniquely identifiable.

What is Spintax?

{ | } *PROJECT
TAILSPIN*

"Spintax"

(sentence permutation syntax)

Writing spintax is fun!

{Writing|Creating} spintax is
{fun|great}!

- > Writing spintax is fun!
- > Creating spintax is fun!
- > Writing spintax is great!
- > Creating spintax is great!

What is Solid Spintax?

Solid Security's advanced standardized spintax format

- Integer Switches
- Nested Switches
- Global Switches
- Ruleset Operations
- Special Characters
- Version Indexing

100 lines (77 sLoc) | 4.6 KB

RawBlameHistory

Solid's Spintax Cheatsheet

This document is intended to serve as a quick reference for the Solid Spintax.

Solid Spintax is an extended and standardized version of the spintax that is used by most spinning software. As such, many of the features outlined below are not supported by all spinners. However, the syntax is designed to be backwards-compatible with existing spintax generators.

This document is not a formal specification — it's merely intended to be used as an easily-digestible overview of what the format can offer.

1. Switches

1.1. String Switch

A string switch is specified using curly braces (`{}`) with options separated by a vertical bar (`|`).

The most basic spintax element is a string switch. The spun text will contain one of the specified options.

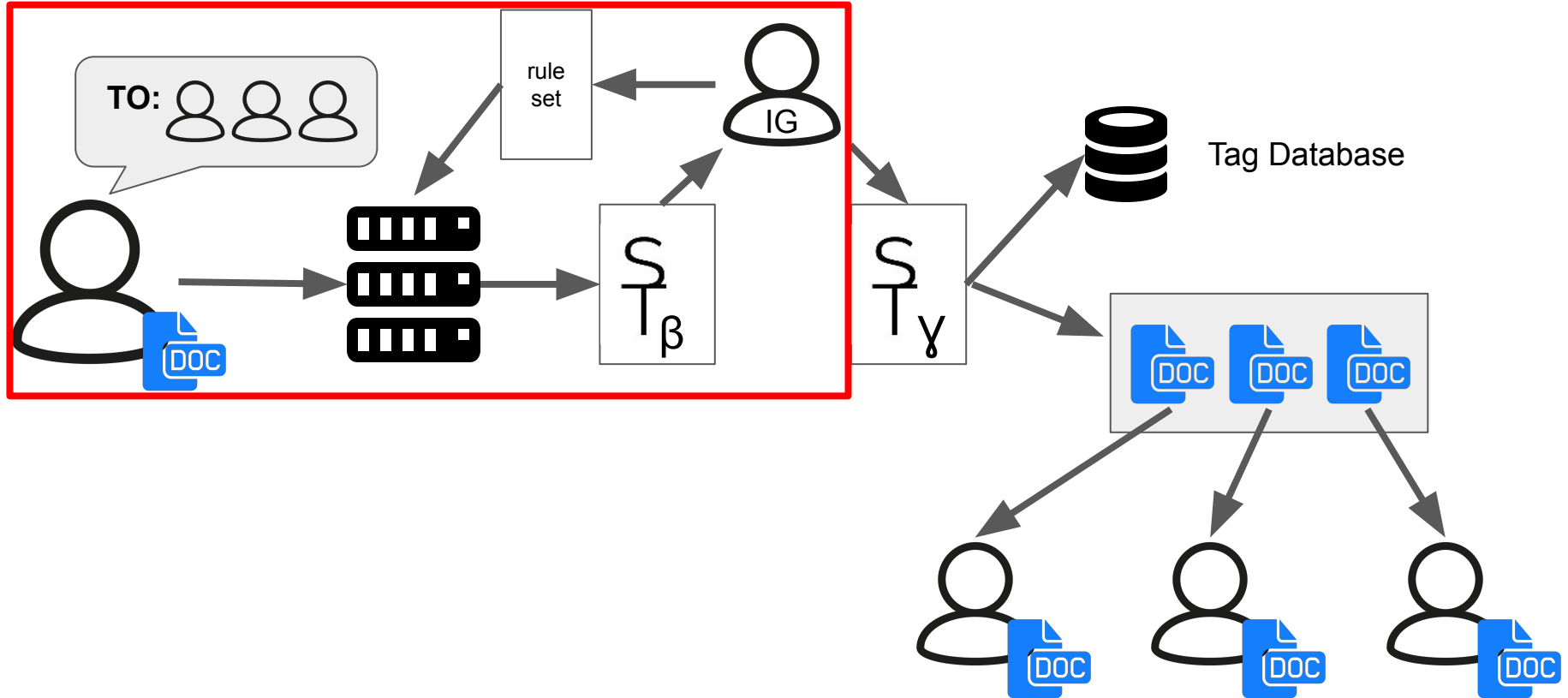
Basic Example:

```
{Hello|Goodbye} World!
```

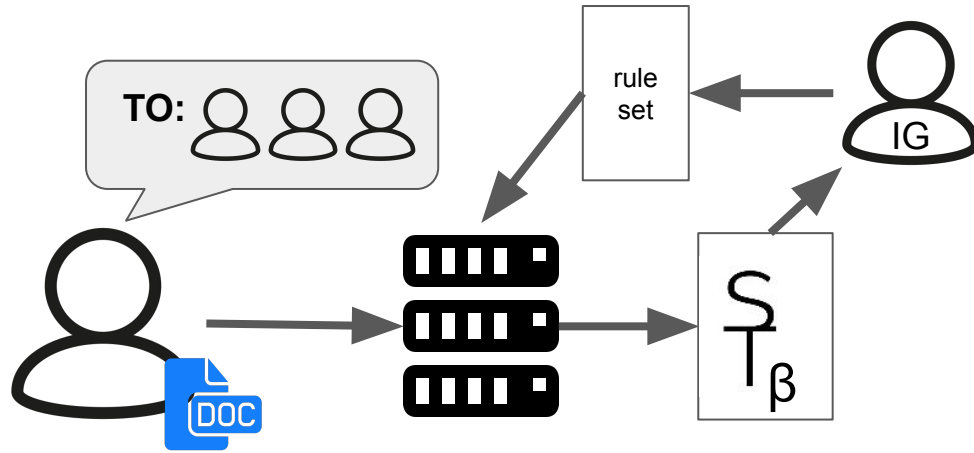
Possible Outputs (2 total):

- "Hello World!"
- "Goodbye World!"

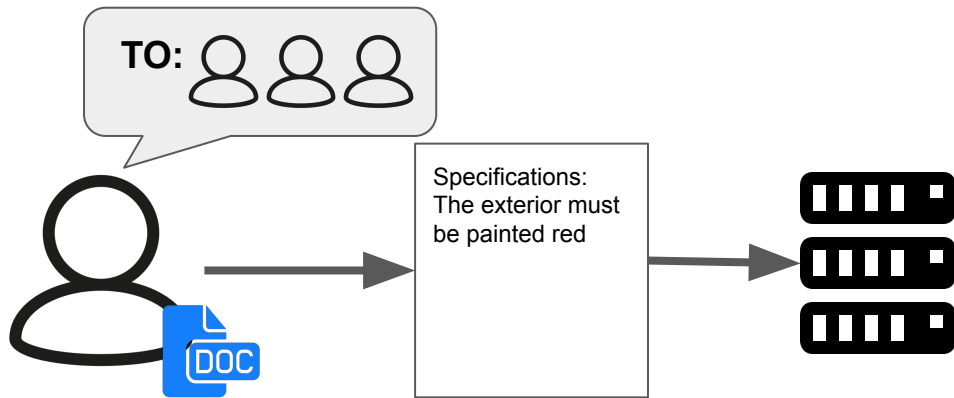
With Tailspin



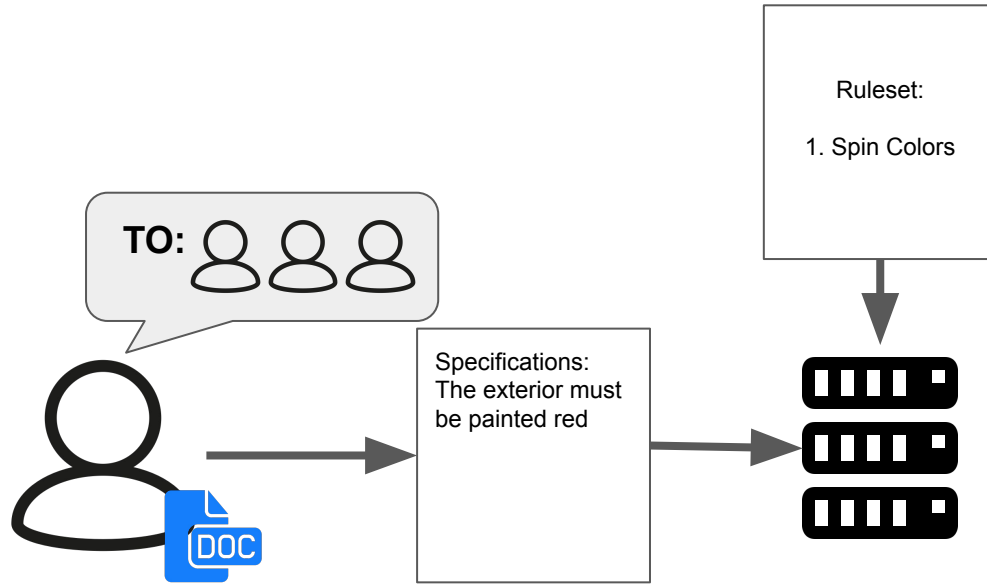
With Tailspin



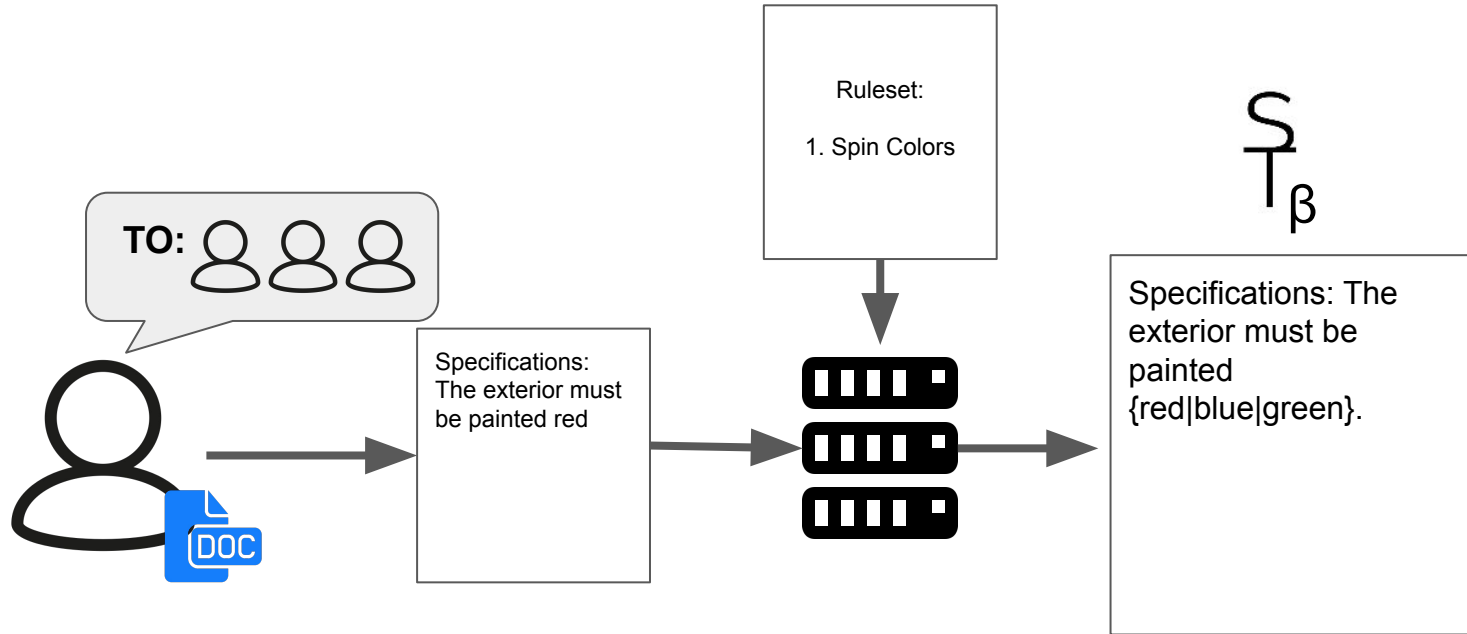
With Tailspin



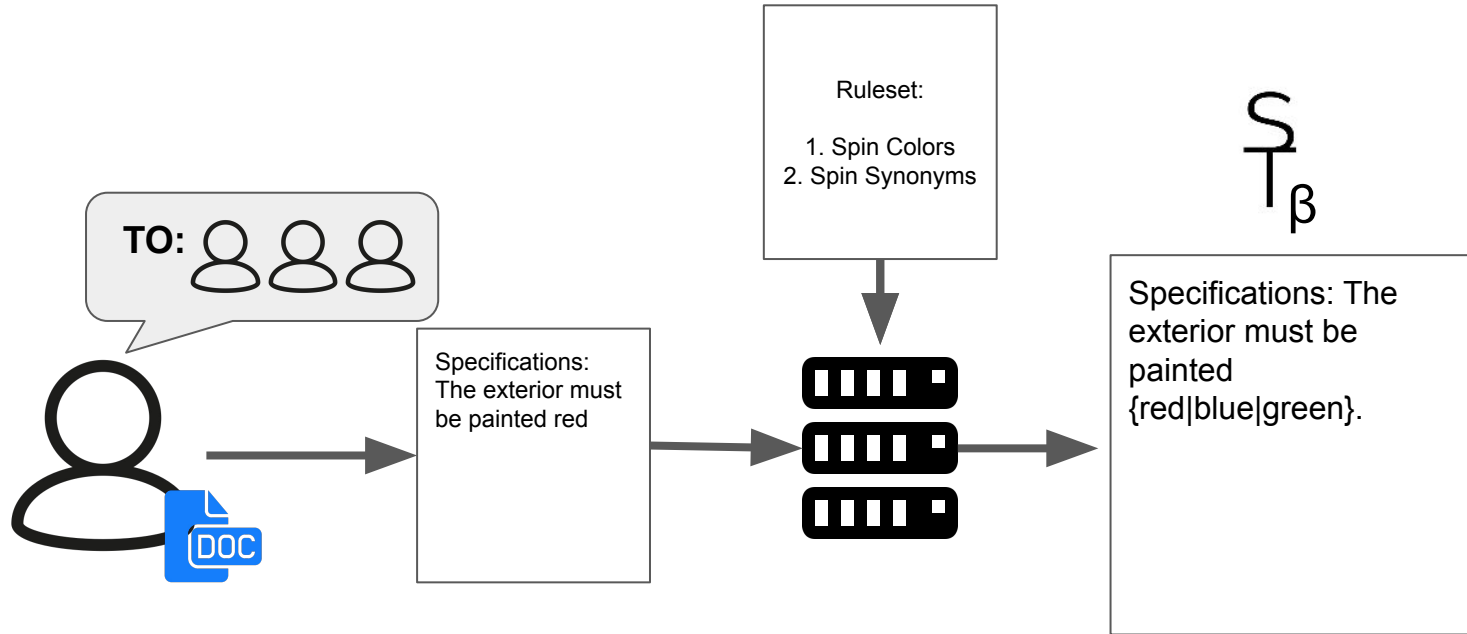
With Tailspin



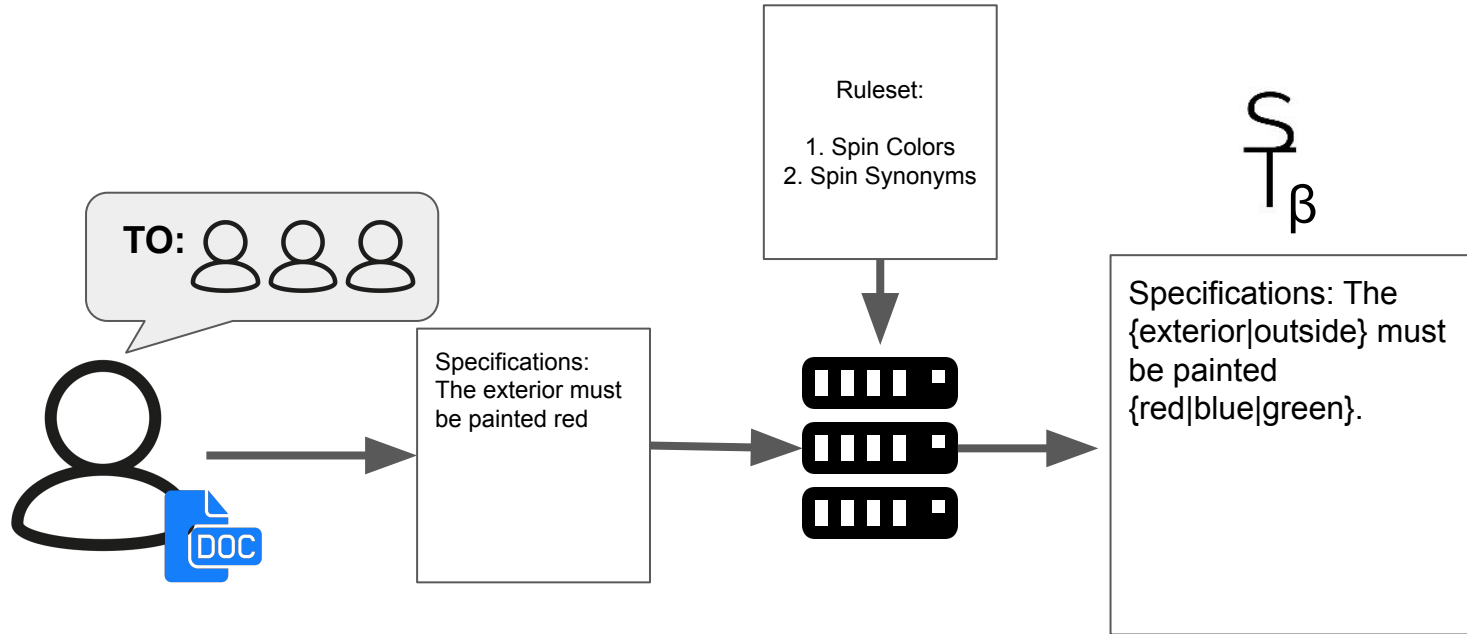
With Tailspin



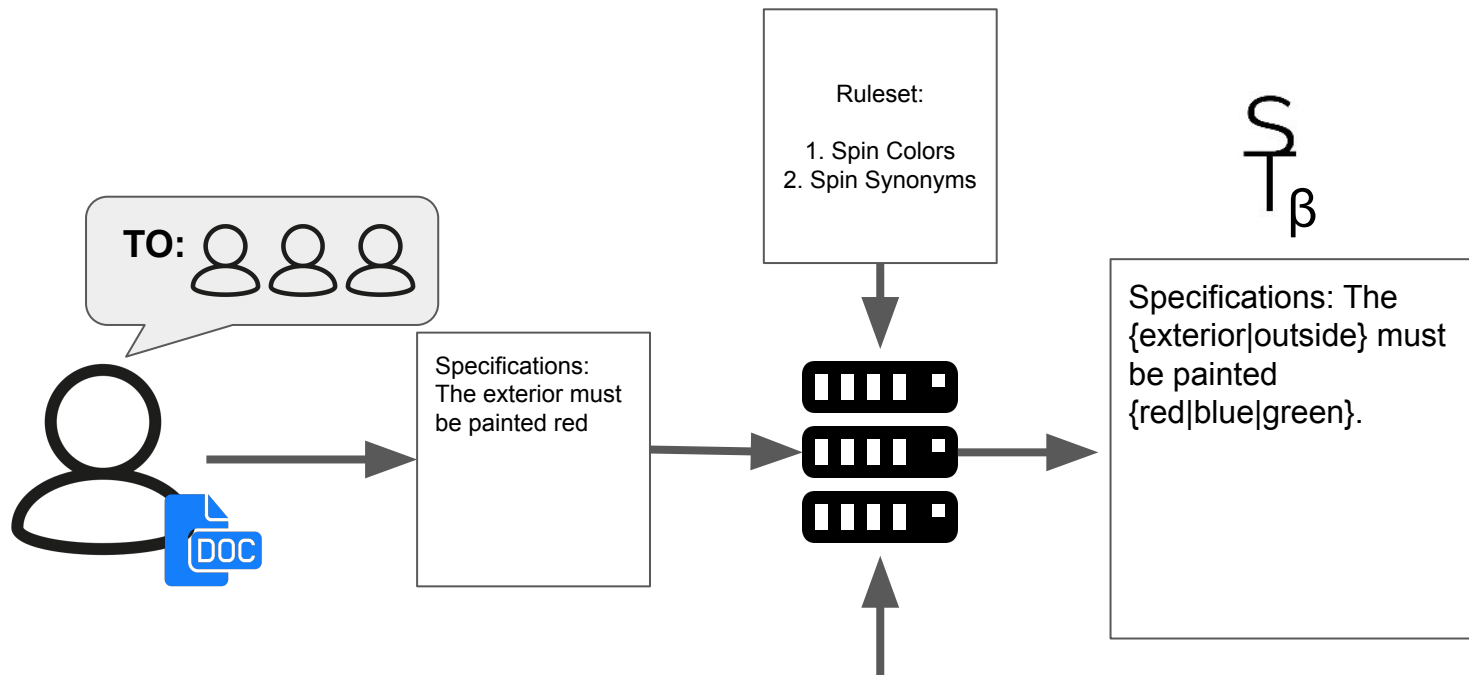
With Tailspin



With Tailspin

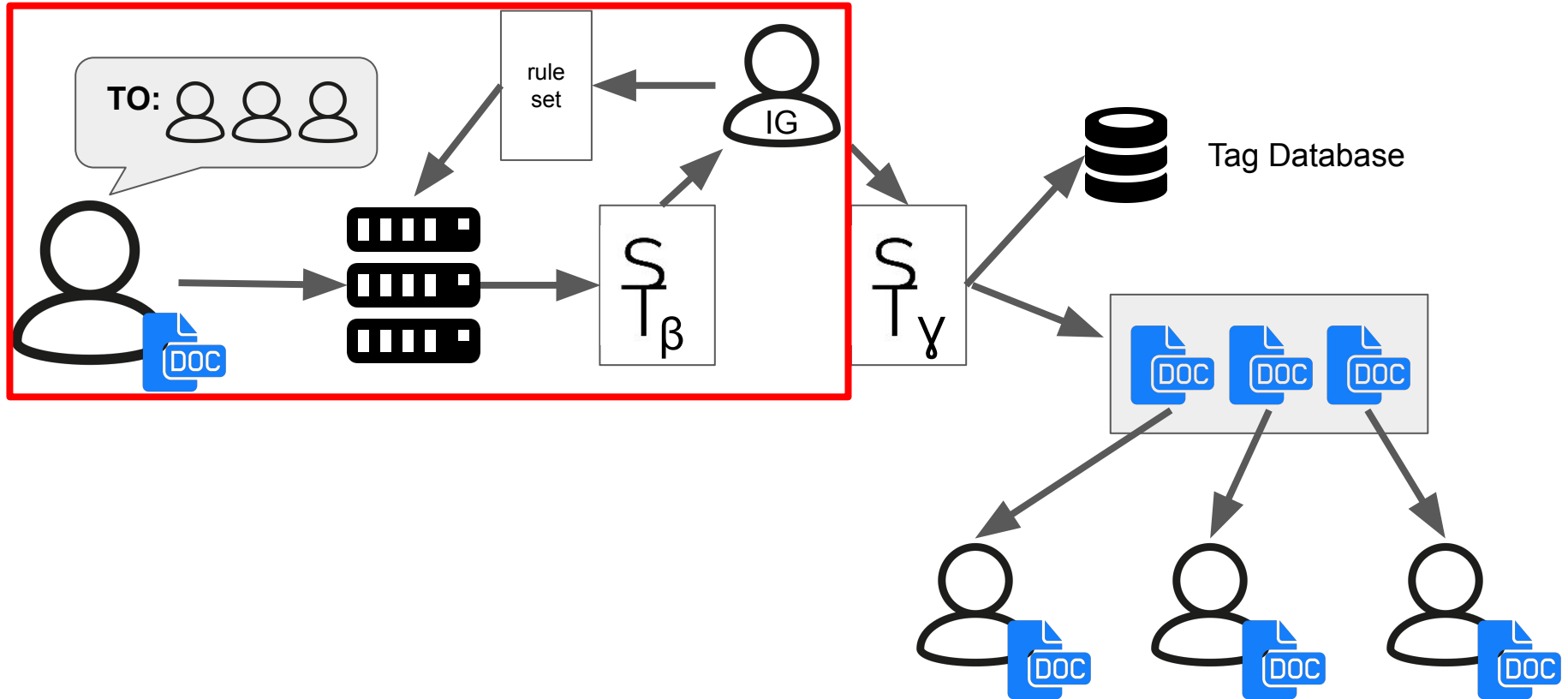


With Tailspin

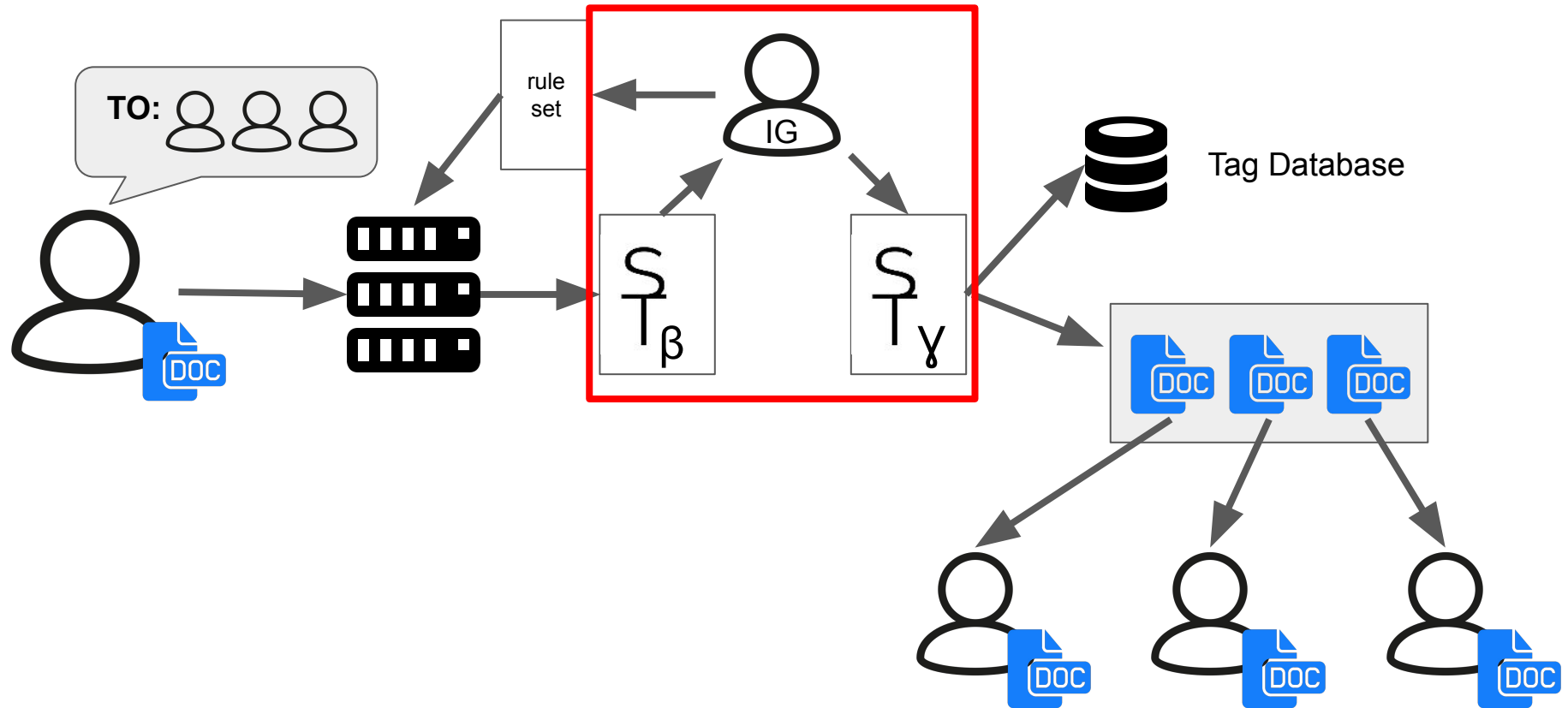


**Doing this convincingly is a
very difficult NLP challenge!**

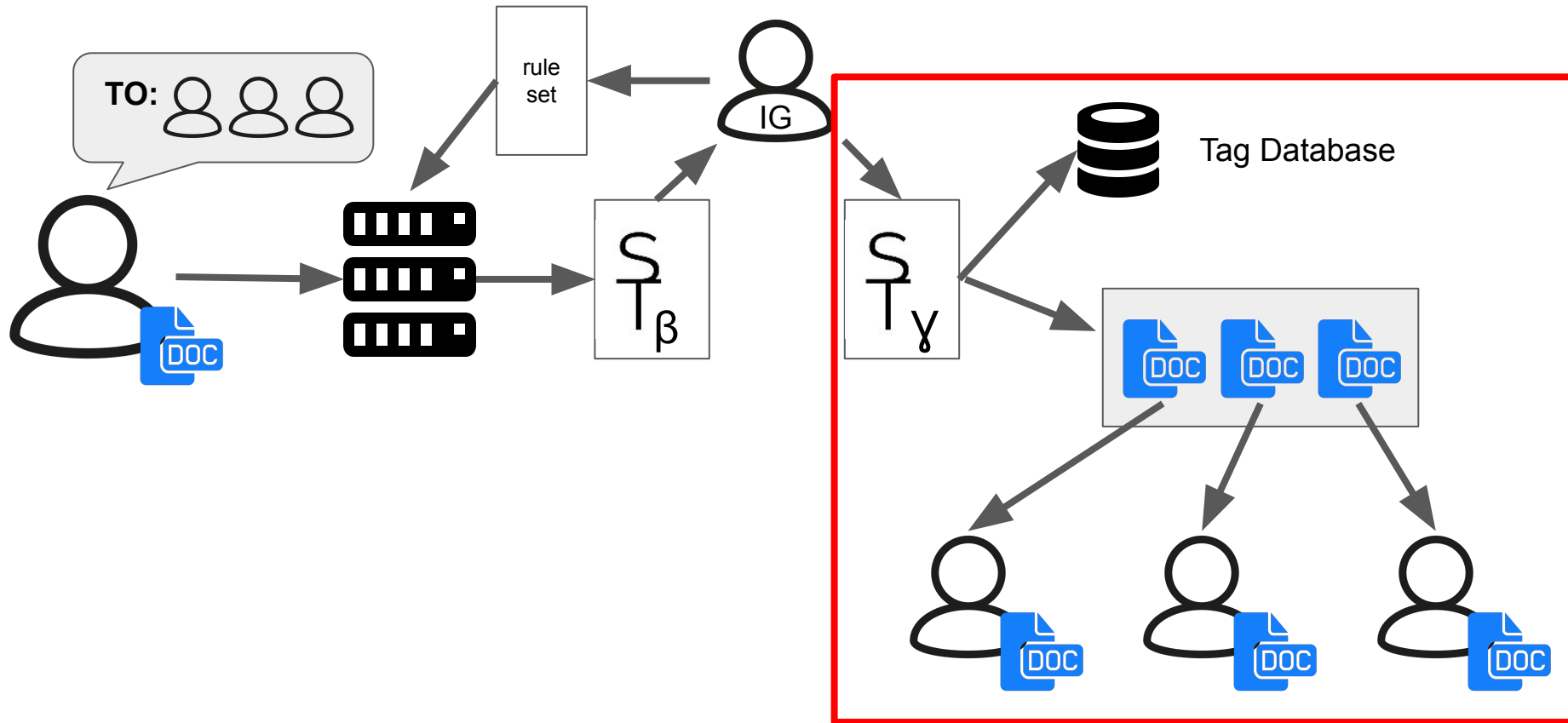
With Tailspin



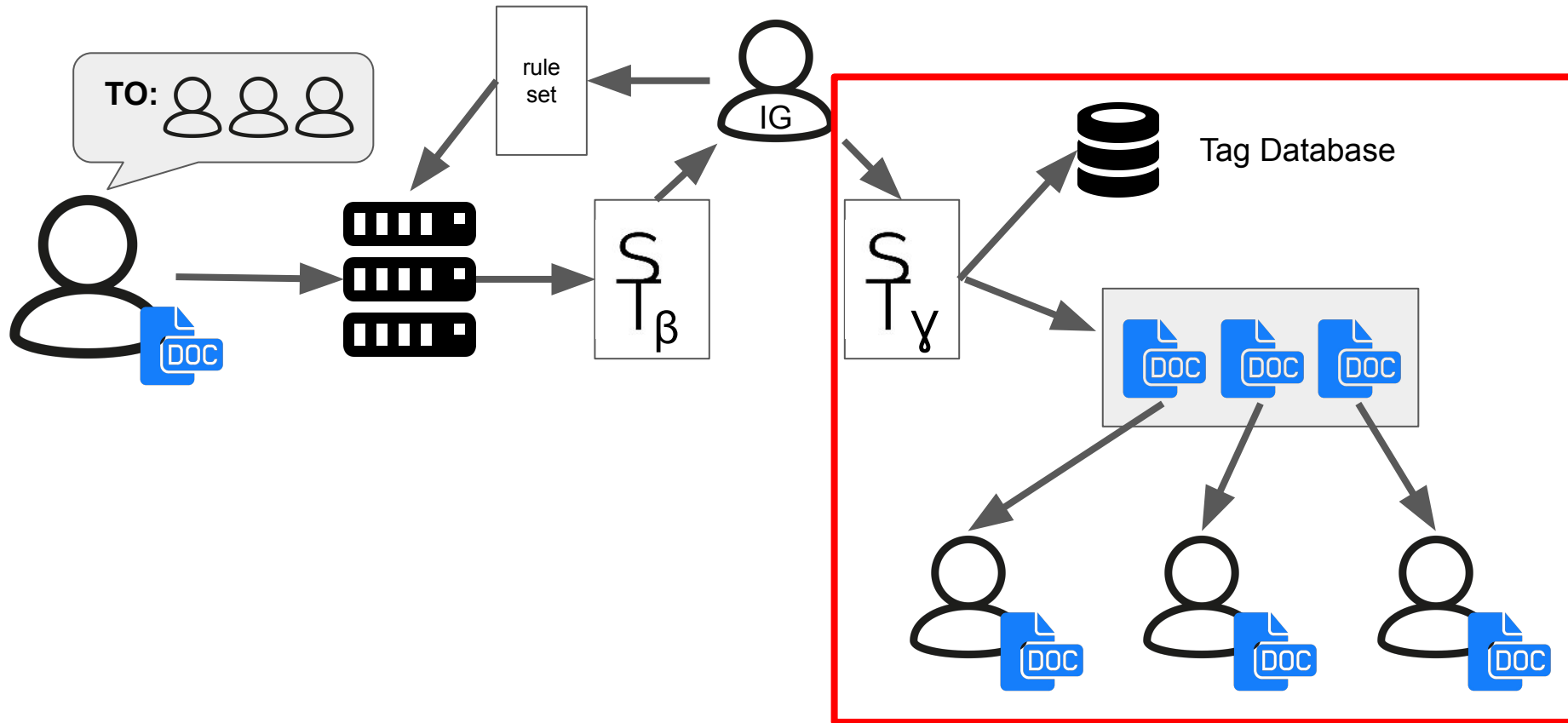
With Tailspin



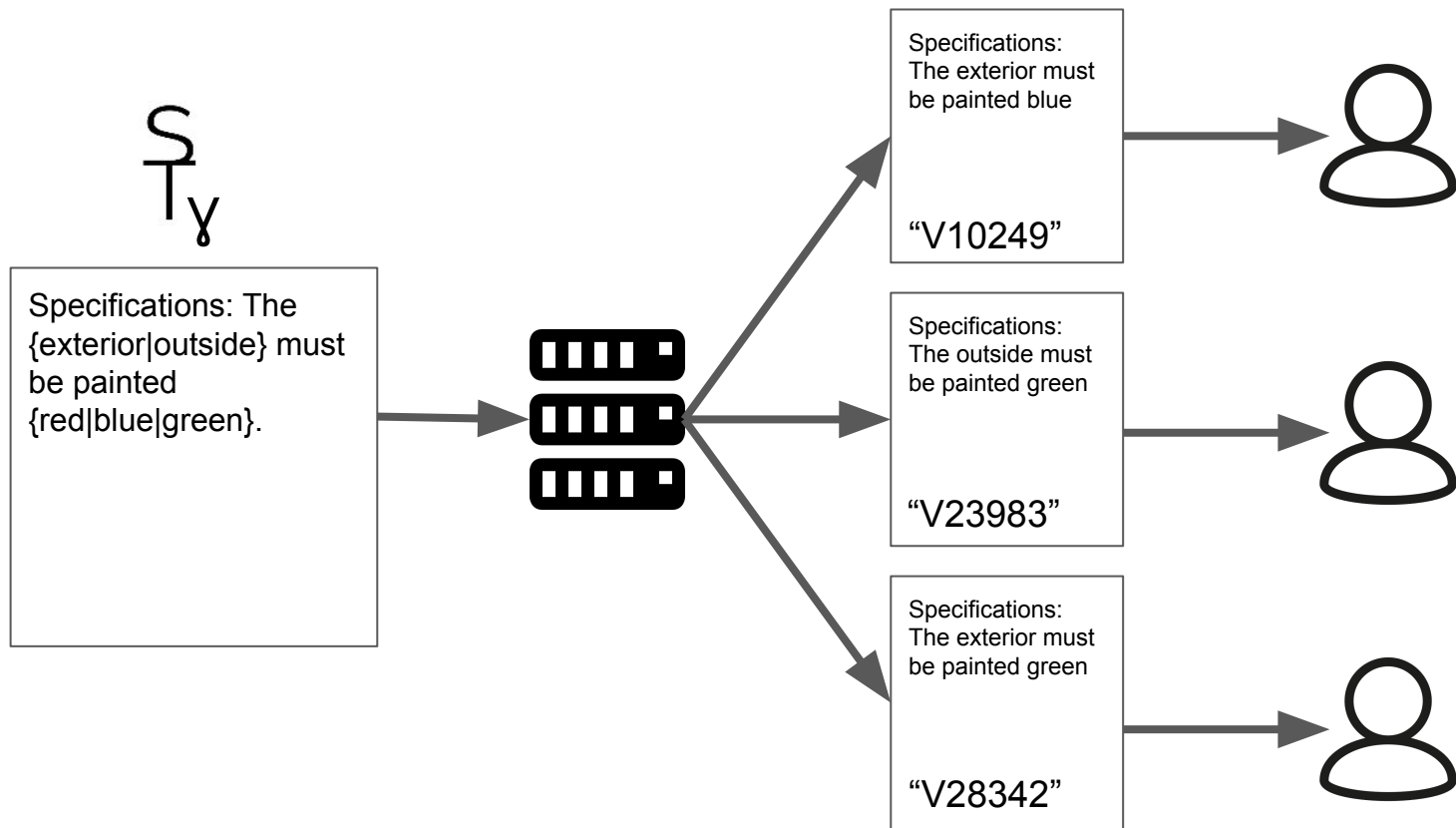
With Tailspin



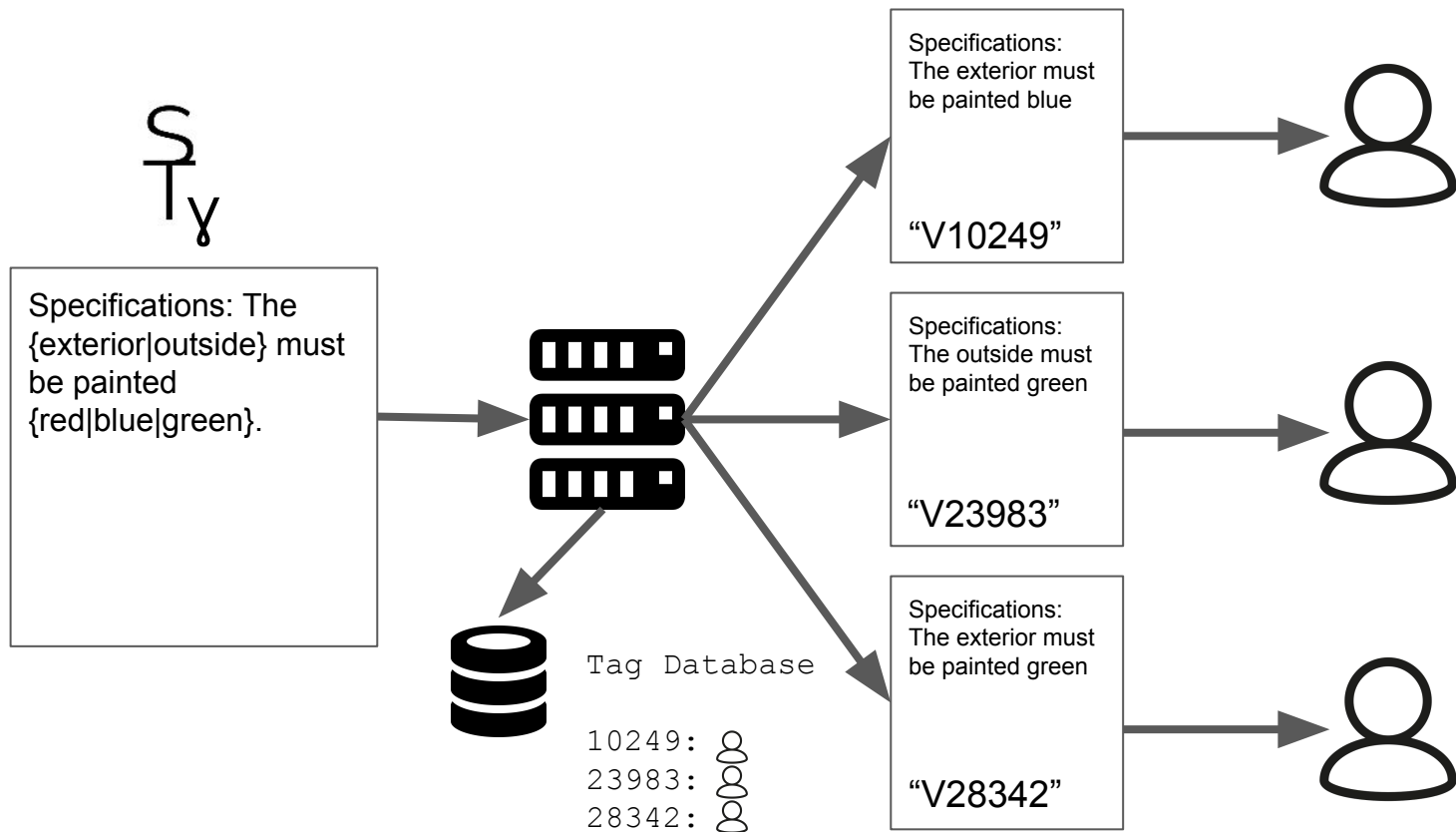
With Tailspin



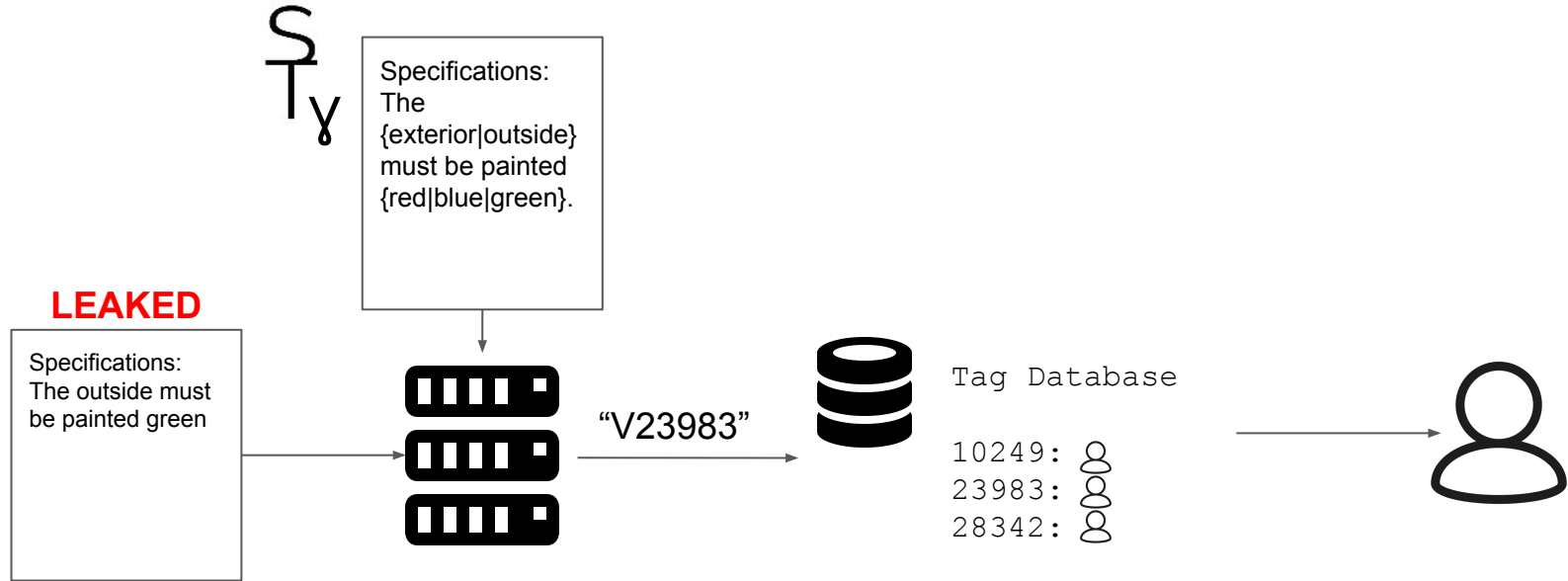
With Tailspin



With Tailspin



With Tailspin





***Objective 2: use data science to
build a dynamic threat model.***

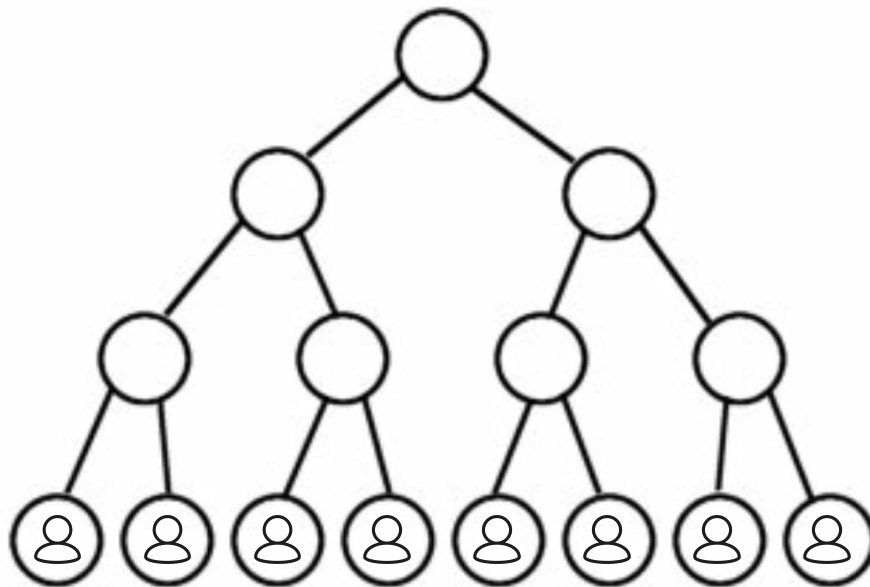
Red-Black Modelling

Red-Black Modelling uses individual data points to construct a comprehensive risk model.

Red-Black Modelling

Red-Black Modelling uses individual data points to construct a comprehensive risk model.

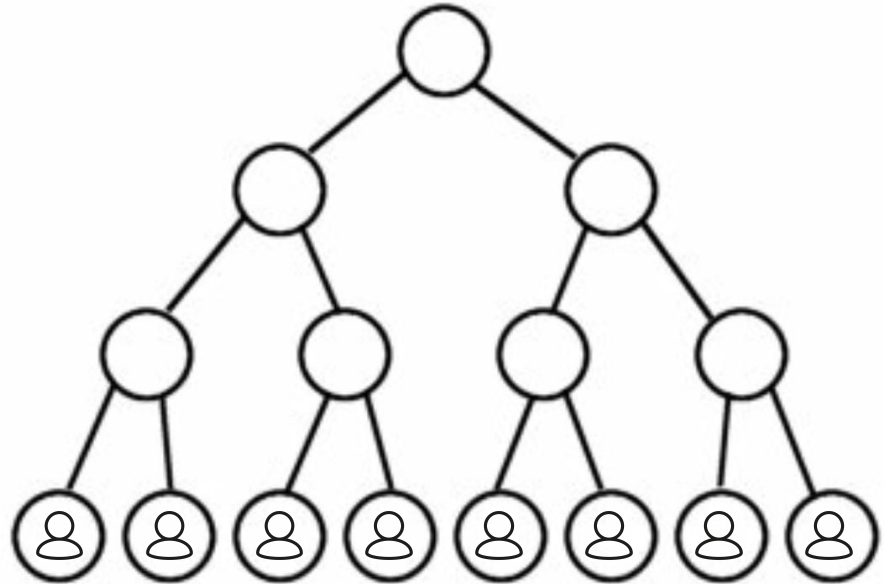
A tree is constructed with document recipients at the leaf nodes.



Red-Black Modelling (Simple Example)

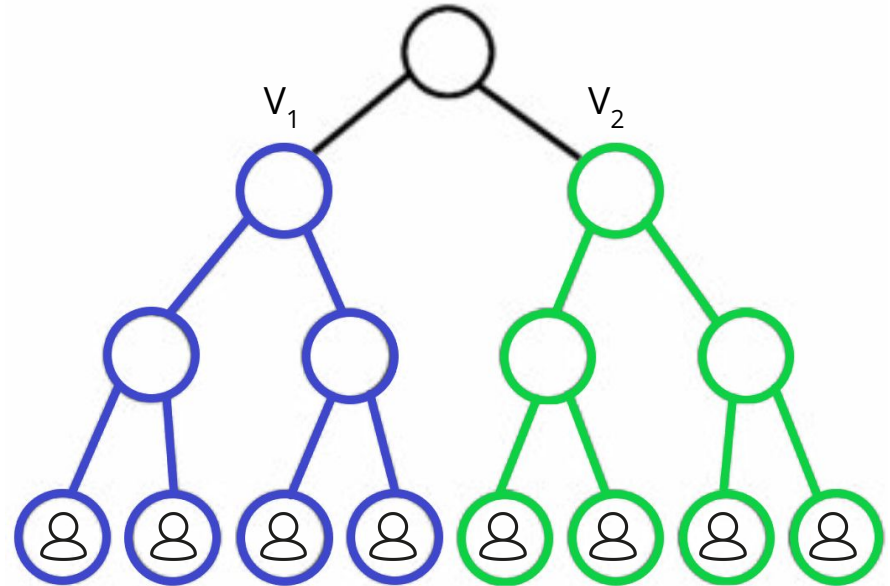
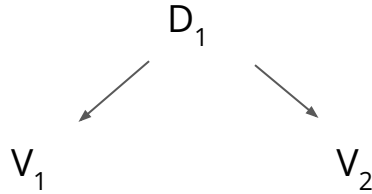
Suppose you have a piece of data you wish to share with a set of personnel...

D_1

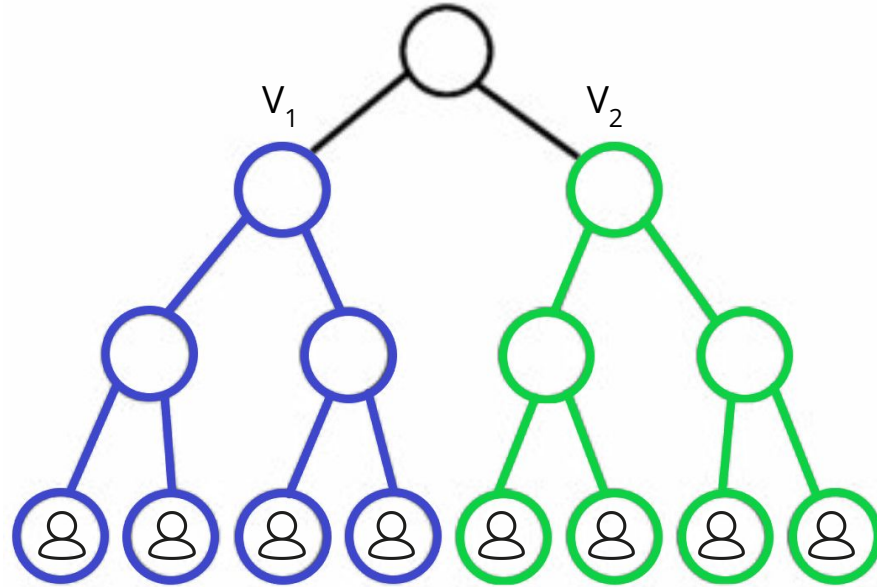


Red-Black Modelling (Simple Example)

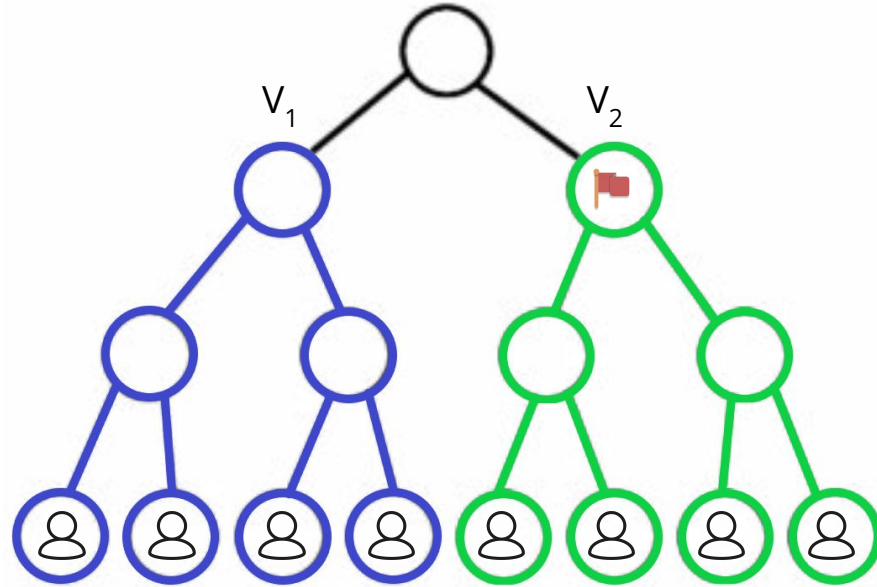
Suppose you have a piece of data you wish to share with a set of personnel...



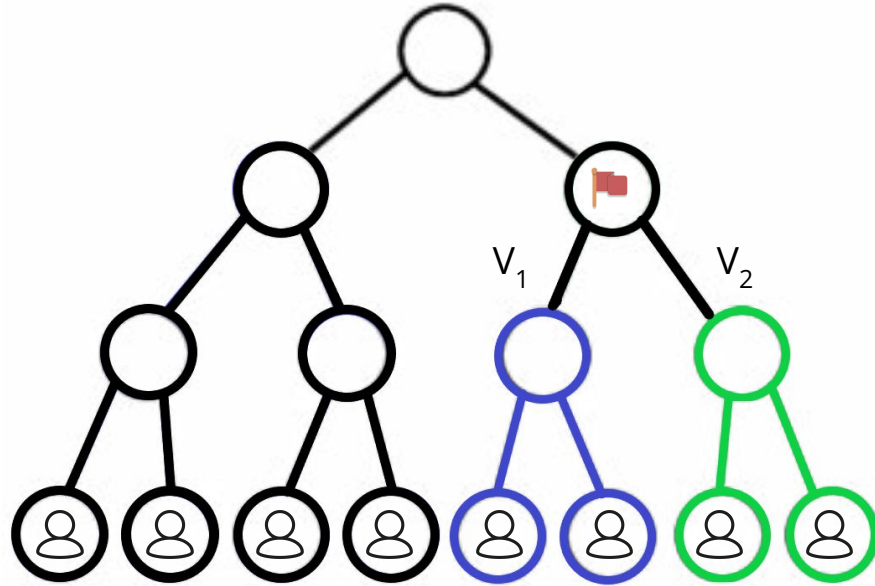
Red-Black Modelling (Simple Example)



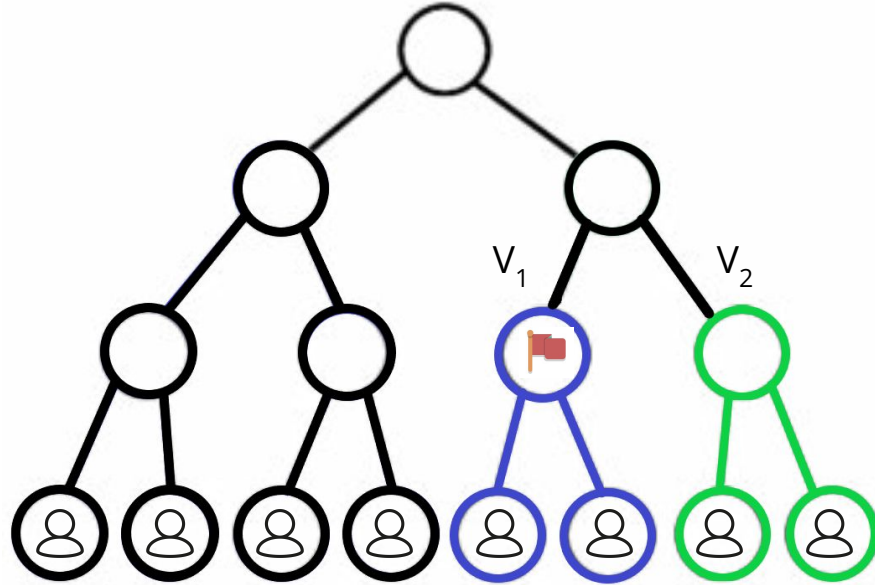
Leak Occurs at v_2



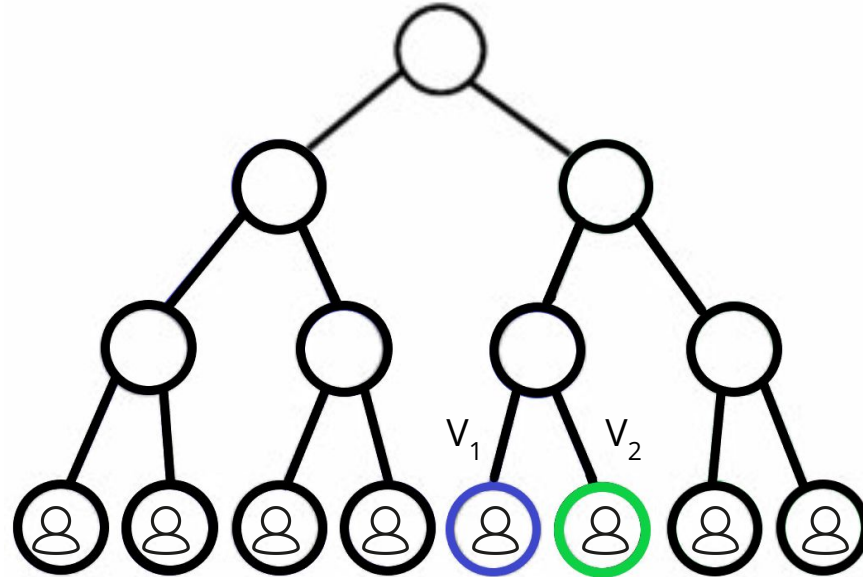
D_2 : v_1 and v_2 sent relative to flag



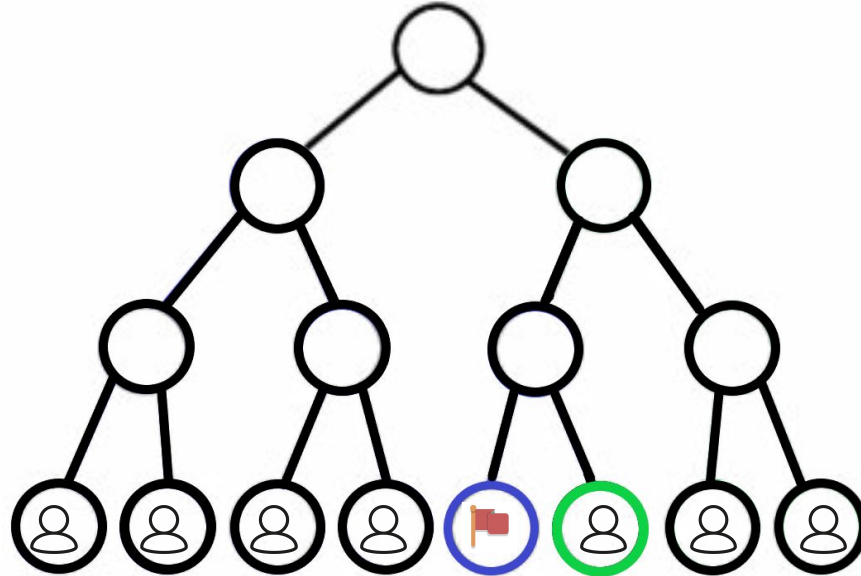
Leak Occurs at v_1



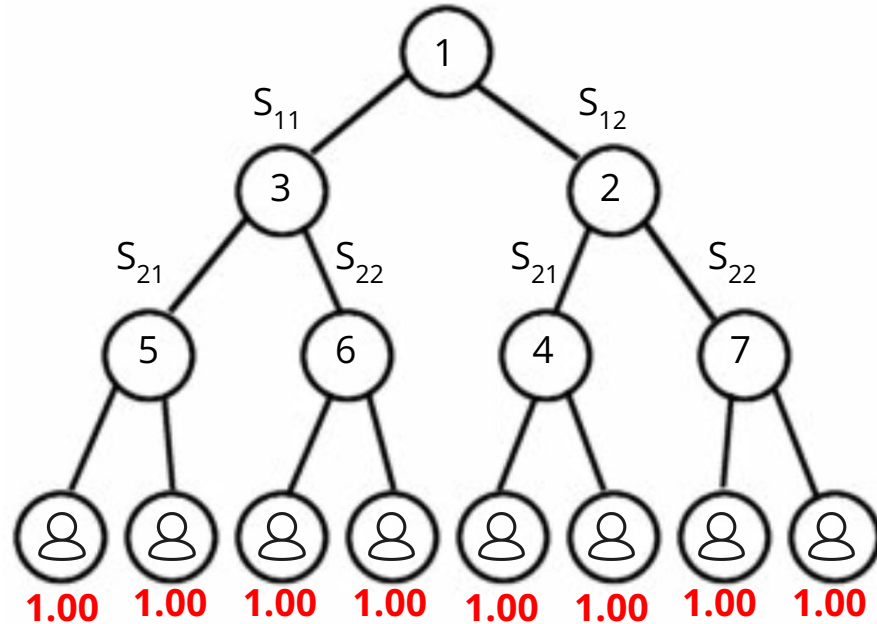
D_3 : v_1 and v_2 sent relative to flag



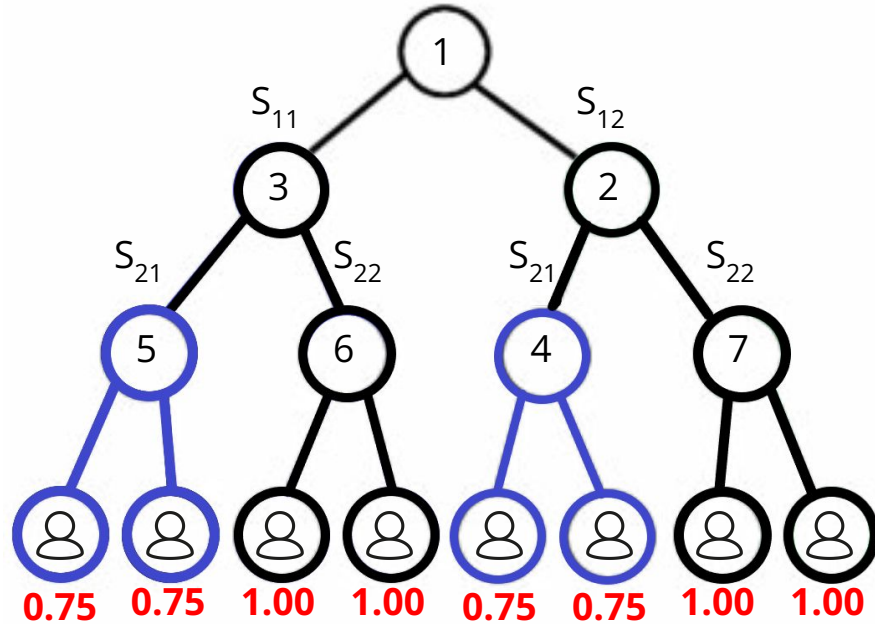
Leak Occurs at v_1 : suspect identified



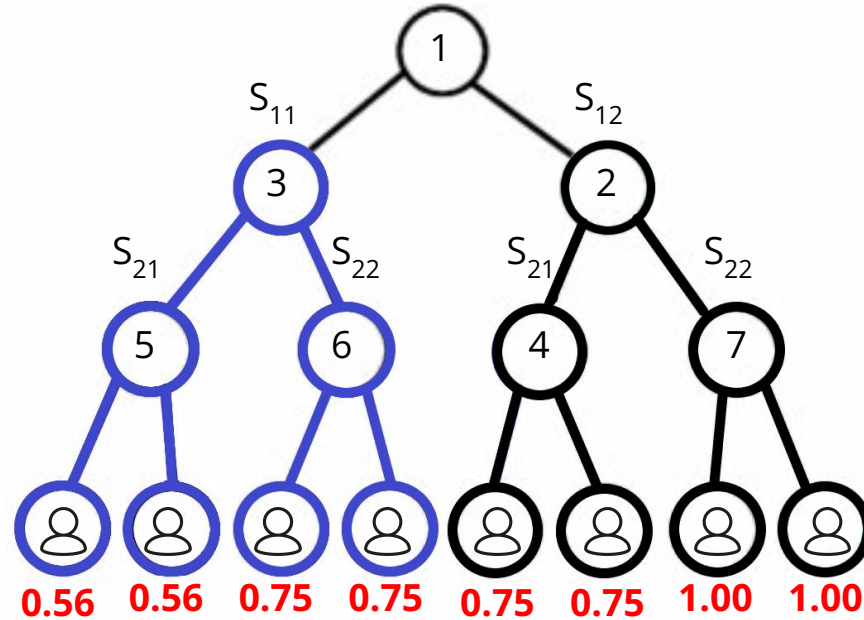
True Red-Black Threat Model



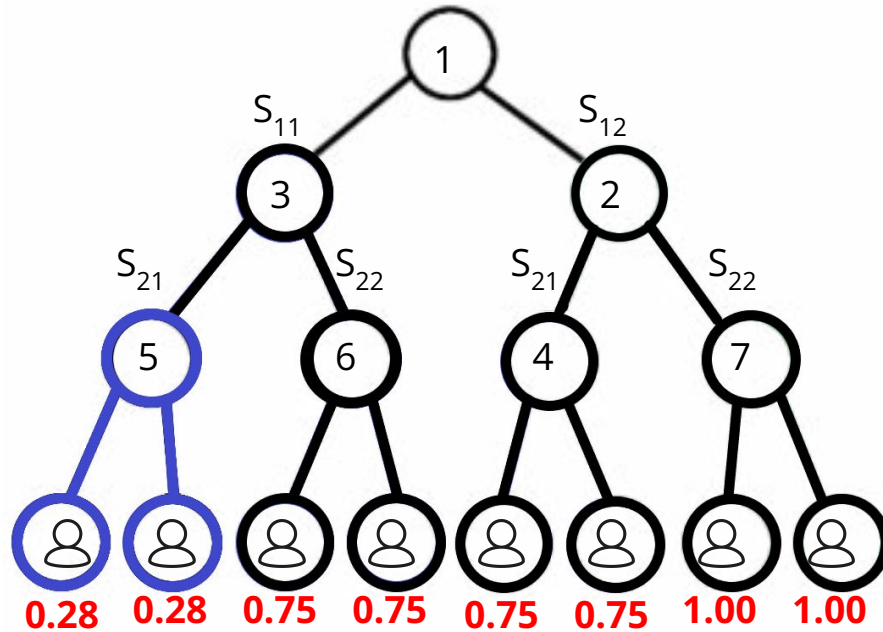
True Red-Black Threat Model



True Red-Black Threat Model



True Red-Black Threat Model



Solid's Spintax Cheatsheet

This document is intended to serve as a quick reference for the Solid Spintax.

Solid Spintax is an extended and standardized version of the spintax that is used by most spinning software. As such, many of the features outlined below are not supported by all spinners. However, the syntax is designed to be backwards-compatible with existing spintax generators.

This document is not a formal specification — it's merely intended to be used as an easily-digestible overview of what the format can offer.

