Fingerprinting Hidden Service Circuits from a Tor Middle Relay

Marc Juarez\textsuperscript{1} Rob Jansen\textsuperscript{2} Rafael Galvez\textsuperscript{1} Tariq Elahi\textsuperscript{1}
Claudia Diaz\textsuperscript{1} Matthew Wright\textsuperscript{3}

\textsuperscript{1}imec-COSIC KU Leuven, \textsuperscript{2}U.S. Naval Research Laboratory, \textsuperscript{3}Rochester Institute of Technology

Motivation

- Kwon et al. propose a decision tree to classify different Tor circuits at the entry.
- However:
  - Decision trees are sensitive to slight modifications of its features.
  - The features may vary at different positions of the circuit.

Research Questions

- Is the attack still effective at a middle?
- Can we use random forests to make it more robust?

Client-RP vs Other

Client-RP is the client-side of a HS circuit.

\begin{itemize}
  \item First 4 cells:
  \begin{itemize}
    \item Client-RP (13)
    \item Other (40)
  \end{itemize}
  \item First 3 cells:
  \begin{itemize}
    \item Client-RP (388)
    \item Other (3959)
  \end{itemize}
  \item # outgoing cells:
  \begin{itemize}
    \item Client-RP
    \item Other
  \end{itemize}
\end{itemize}

Features:

- First \( n \) cells during circuit construction
- Total number of in/out cells
- Total duration time

Results: This decision tree achieves 100\% accuracy on traffic traces collected from a Tor middle node.

Adversary Model

Goal: distinguish .onion from www

Capabilities:

- Owns a middle: low cost
- Passive
- Records Tor cells

Random Forest Attack

- Random forests: get an answer from multiple decision trees created by randomizing features and training data.
- Results: high success rates at both the client and the middle.
- Random forests generalize better and thus are a more robust model for the attack than decision trees.

Conclusions and Future Work

- The attack is as accurate at the middle as at the entry.
- The random-forest-based attack is more robust than using a single decision tree, and thus, it should be used to evaluate defenses against circuit fingerprinting.
- We will design new defenses against traffic fingerprinting in Tor (from any position in Tor), and evaluate them using the random forest-based attack.