DE-ANONYMIZATION IN BITCOIN

GEORGE KAPPOS, HAAROON YOUSAF, RAINER STÜTZ,
SOFIA ROLLET, BERNHARD HASLHOFER, AND
SARAH MEIKLEJOHN

BITCOIN IS NOT ANONYMOUS

Quantitative Analysis of the Full Bitcoin Transaction Graph

Dorit Ron and Adi Shamir

An Analysis of Anonymity in the Bitcoin System

Fergal Reid and Martin Harrigan

BitIodine: Extracting Intelligence from the Bitcoin Network

Michele Spagnuolo, Federico Maggi, and Stefano Zanero

Evaluating User Privacy in Bitcoin

Elli Androulaki[1], Ghassan O. Karame[2], Marc Roeschlin[1], Tobias Scherer[1], and Srdjan Capkun[1]

A Fistful of Bitcoins: Characterizing Payments Among Men with No Names

Sarah Meiklejohn Marjori Pomarole Grant Jordan
Kirill Levchenko Damon McCoy[†] Geoffrey M. Voelker Stefan Savage

BITCOIN IS NOT ANONYMOUS

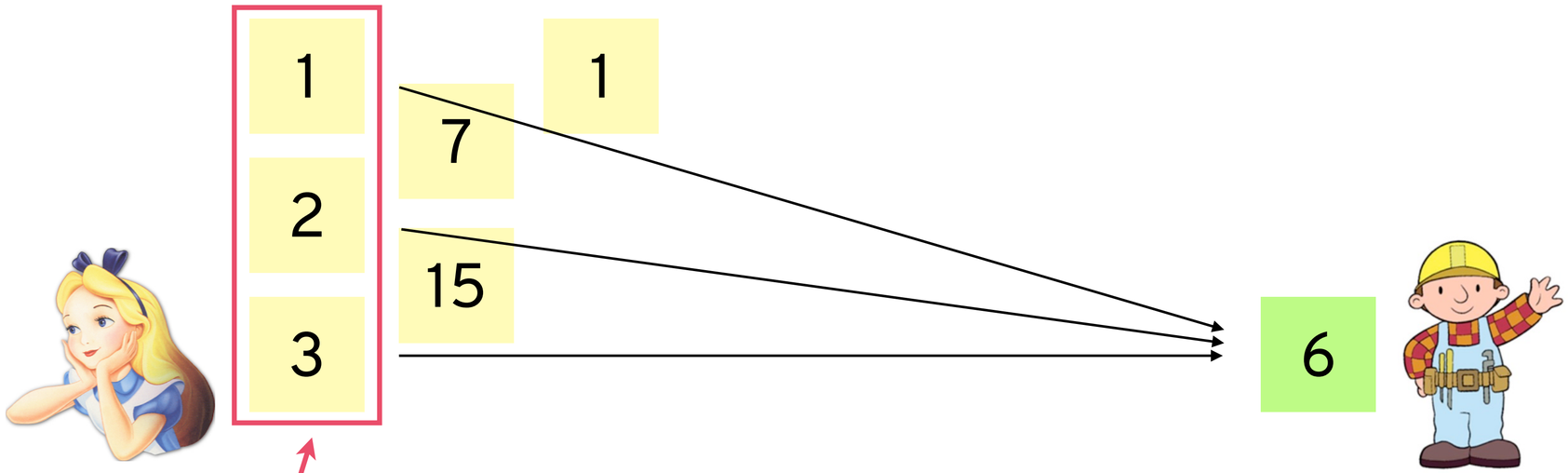
Prosecutors Trace $13.4M in Bitcoins From the Silk Road to Ulbricht's Laptop

Global Disruption of Three Terror Finance Cyber-Enabled Campaigns

US Officials Arrest Alleged Operator of $336M Bitcoin Mixing Service

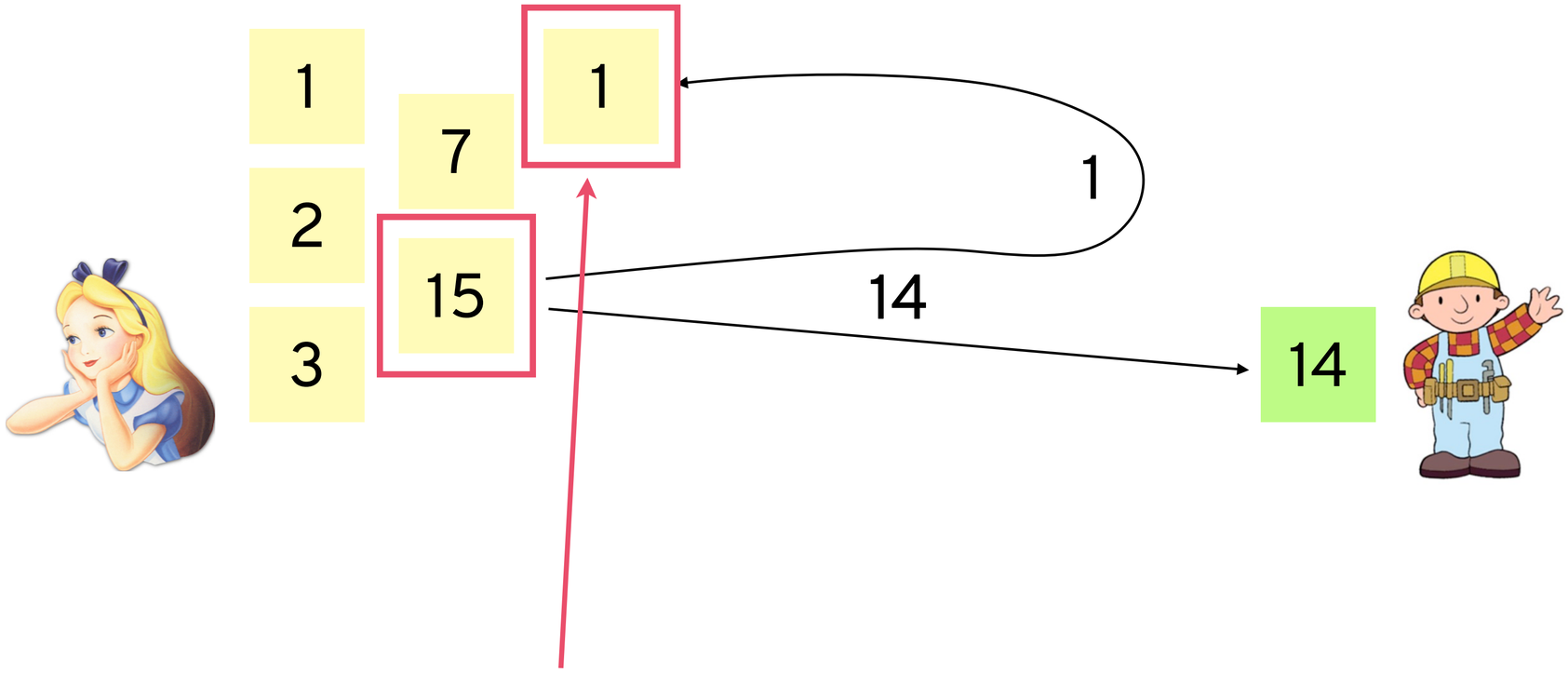
Inside the Bitcoin Bust That Took Down the Web's Biggest Child Abuse Site

CLUSTERING BY INPUT



multi-input / co-spend heuristic: the same entity controls these addresses

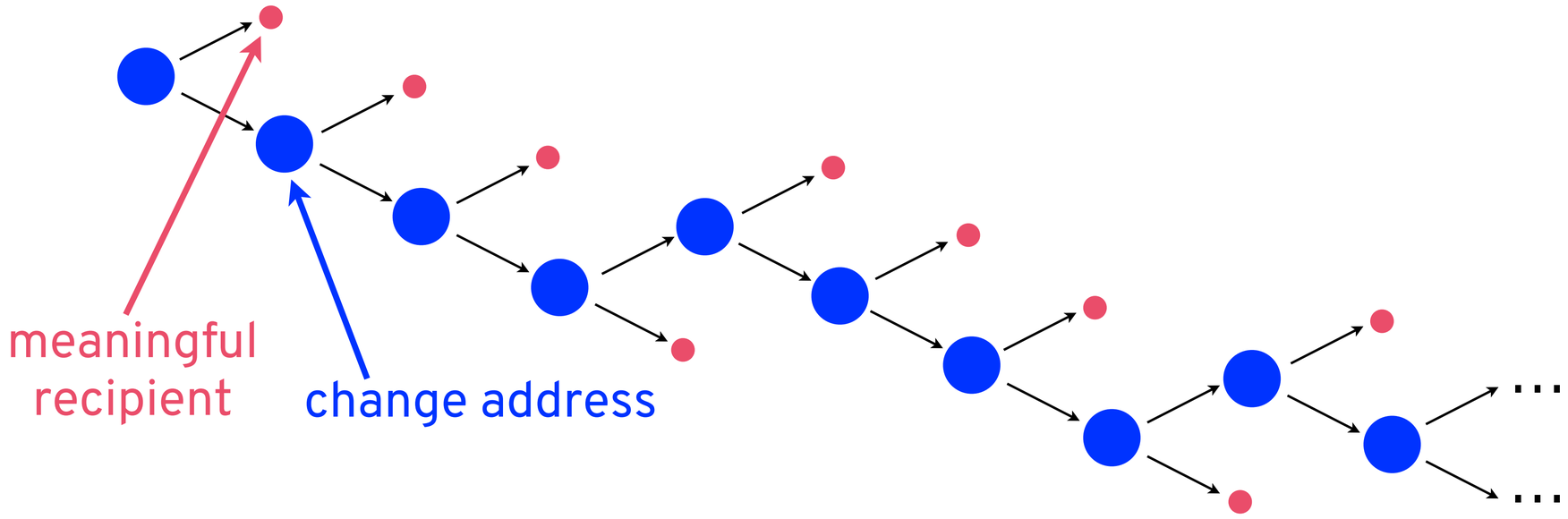
CLUSTERING BY CHANGE



change heuristic: the input entity also controls the change address

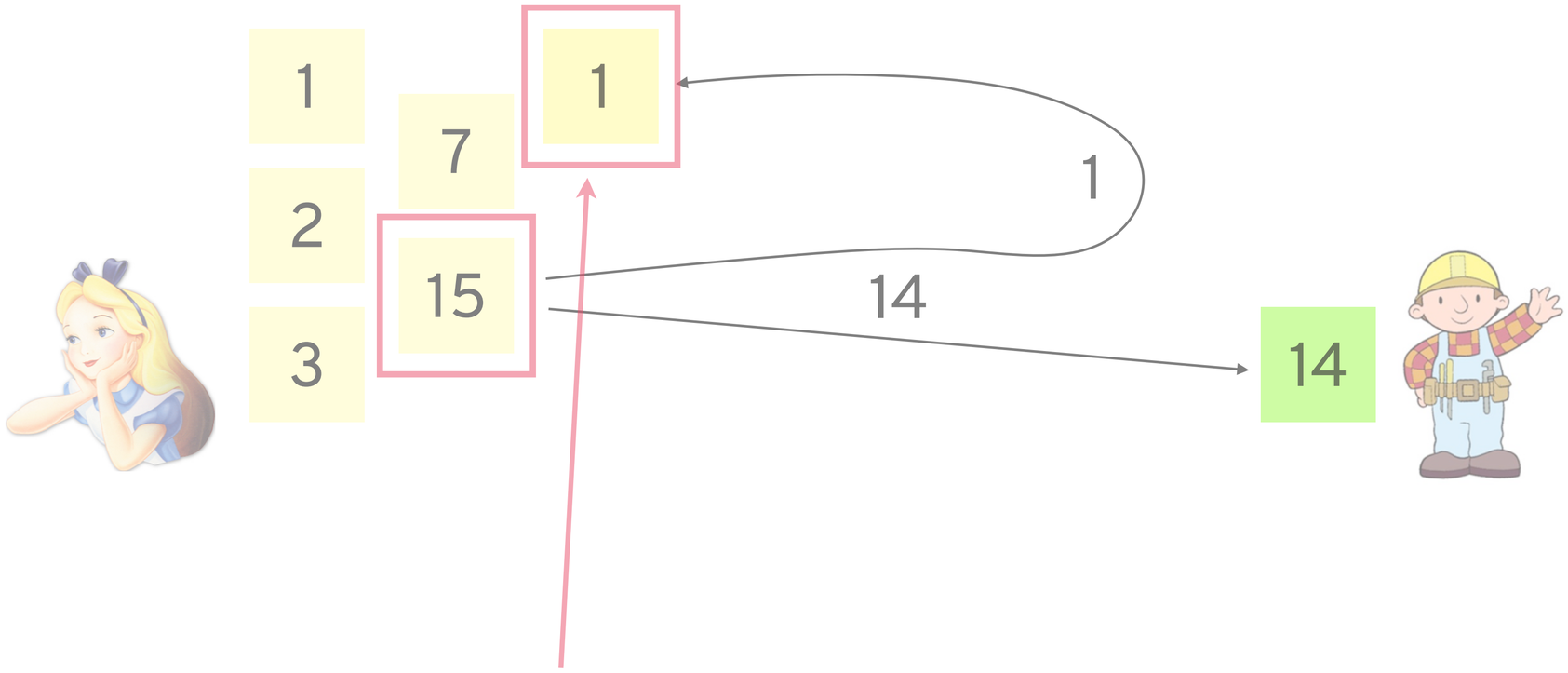
FOLLOWING BITCOINS

Identifying change addresses also allows us to see when bitcoins meaningfully change hands, and thus follow **peel chains**



Identifying recipients of these “peels” potentially de-anonymizes user

CLUSTERING BY CHANGE



change heuristic: the input entity also controls the change address

HOW DO WE ACTUALLY IDENTIFY THE CHANGE ADDRESS?

DIVERSE FEATURES

The Bitcoin protocol has changed a fair amount since 2013!

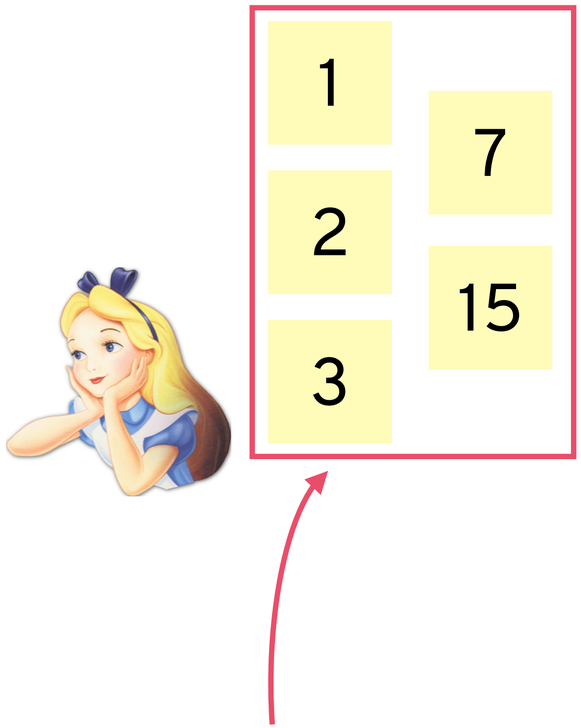
New **transaction features**: locktime, RBF, etc. [MN22]

New **address types**: P2SH, Bech32, etc.

Can also define the **change strategy** of a multi-input cluster according to where its change addresses are in the list of outputs

- 0: always first
- -1: always last
- 1: always first or last
- none

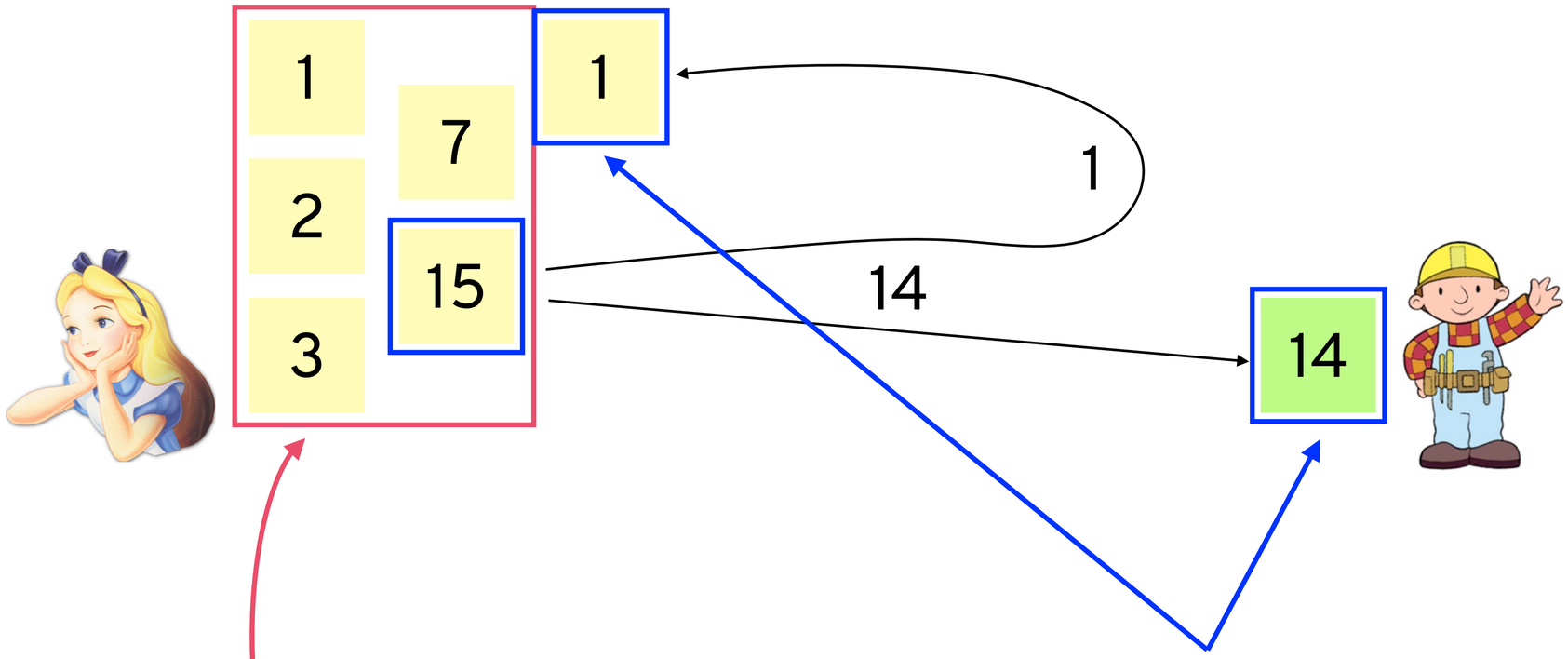
A NEW CHANGE HEURISTIC



for a multi-input cluster, define:

- set of transaction features
- set of address types
- change strategy

A NEW CHANGE HEURISTIC



for a multi-input cluster, define:
- set of transaction features
- set of address types
- change strategy

label as change the unique output address that matches these features (considering the transaction in which it spends its contents)

EVALUATING OUR HEURISTIC

For ground-truth multi-input clusters (C_{addr}, C_{tx}) curated from data provided by Chainalysis, followed peel chains starting at each tx in C_{tx}

Consider two factors

- **expansion rate**: the ratio of new to old transactions
- **false discovery rate**: the ratio of false positives (as identified by Chainalysis tags) to true / unknown positives

EVALUATING OUR HEURISTIC

For ground-truth multi-input clusters (C_{addr}, C_{tx}) curated from data provided by Chainalysis, followed peel chains starting at each tx in C_{tx}

HEURISTIC	EXPANSION	FDR
[AKR+13]	93.03	64.19
[MKJ+13]	79.94	51.64
[GKRN18]	73.7	48.7
[EPY17]	28.6	12.7
[KYS+22]*	124.46	0.02

CONCLUSIONS

Based on (limited) ground-truth data, our change heuristic seems effective in expanding multi-input clusters

Can also be used to validate the results of the multi-input heuristic

Possible to evade by randomizing features

Bitcoin is not anonymous!

THANKS!

ANY QUESTIONS?