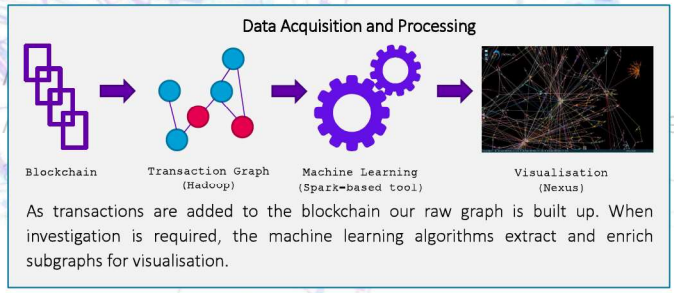
Bitcoin Transaction Analysis with Machine Learning

Jonathan Francis Roscoe <jonathan.roscoe@bt.com>, Jonty Page, Alex Healing

BT Research & Innovation

Background

Although the blockchain does not provide a link to real-world identities, it contains detailed information of all transactions. We employ state-of-the-art machine learning and visualisation to help analysts identify connected wallets.

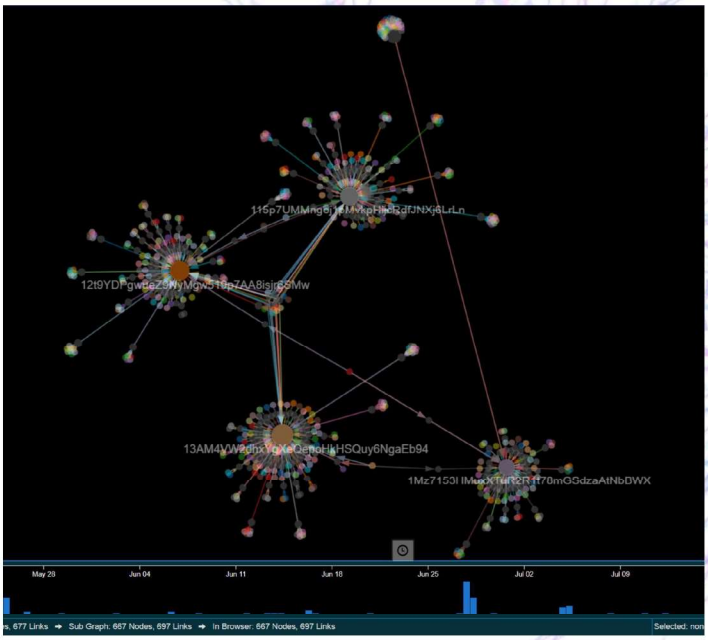


As transactions are added to the blockchain our raw graph is built up. When investigation is required, the machine learning algorithms extract and enrich subgraphs for visualisation.

Machine Learning on the Blockchain

Bitcoin provides a dense dataset that currently lacks purposeful visualisation. Particularly in the absence of large scale equipment[1], data-reduction and abstraction are required for any reasonable visualisation.

Our Spark[2] based platform empowers analysts with access to classifiers and clustering, as well as basic aggregation and filtering. We provide enrichment and automatic analysis to present pertinent information whilst retaining data fidelity.



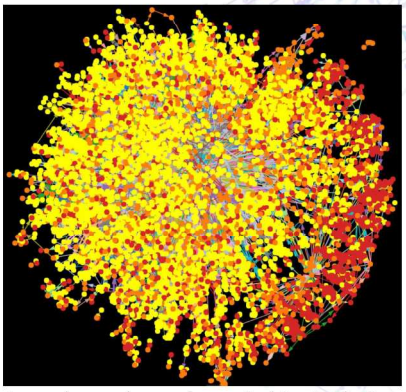
Our Nexus visualisation tool allows us to explored relationships between data points and monitor graphs in realtime. In this example, we are looking at transactions linking ransomware (WannaCry and NotPetya) addresses.

Meiklejohn et al.[3] demonstrate aggregation can be easily achieved and observe variation in graph structure for different services and users. Our research made use of connectivity, frequency and timing for clustering.

Identifying Linked Wallets

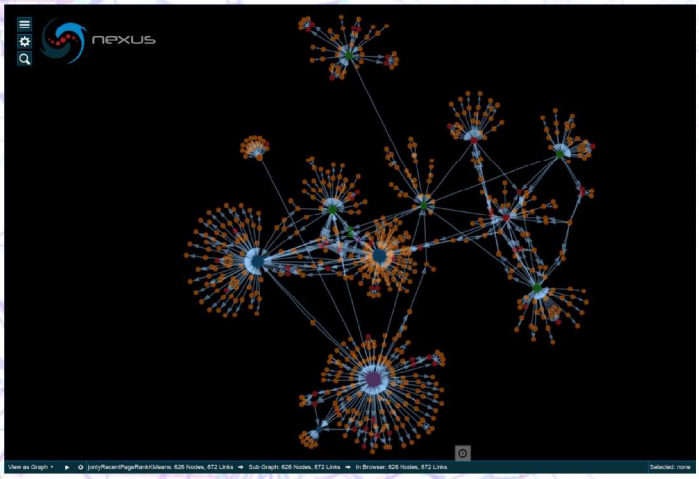
During a recent incident a wallet attributed to a dark web vendor of hacking tools was identified.

We collected linked addresses to 4 degrees, approx. 50,000 nodes.



An overly complex graph, unsuited to visualisation. Constructed from addresses connected to three degrees.

Cluster groups based on transaction frequency and authority/hub scores were used as parameters to visualisation.



A graph of suspicious activity following enrichment with cluster groups based on several node features.

This visualisation provided analysts with new insights and has already aided criminal investigations and future operational trials will research automatic classification of activity at scale.

References

1. **Visualizing Dynamic Bitcoin Transaction Patterns**, McGinn et al., Big Data, vol. 4, 2016
2. **Apache Spark**: https://spark.apache.org/
3. **A Fistful of Bitcoins: Characterizing Payments Among Men with No Names**, Meiklejohn et al., IMC Proceedings, 2013