

# Pattern-based Features

A Data Transformation Pattern

<http://research.microsoft.com/en-us/projects/tark/>

Venkatesh-Prasad Ranganath

Jithin Thomas

Microsoft Research, India

<http://research.microsoft.com/ICSE2013>

# Context / Scenarios

- Compatibility Testing
- Test Prioritization / Test Suite Minimization
- Representative Identification
- Similar Case Recommendation
- Anomaly Detection

# Constraints

## Input Characteristics

- Data is sequential
- Data is structured
- Fields may be irrelevant
- Values may be irrelevant
- Value flow may be relevant

## Output Constraints

- Usable with existing DM/ML algorithms
- Amenable to simple reasoning
- Accessible
- Possess explanatory power

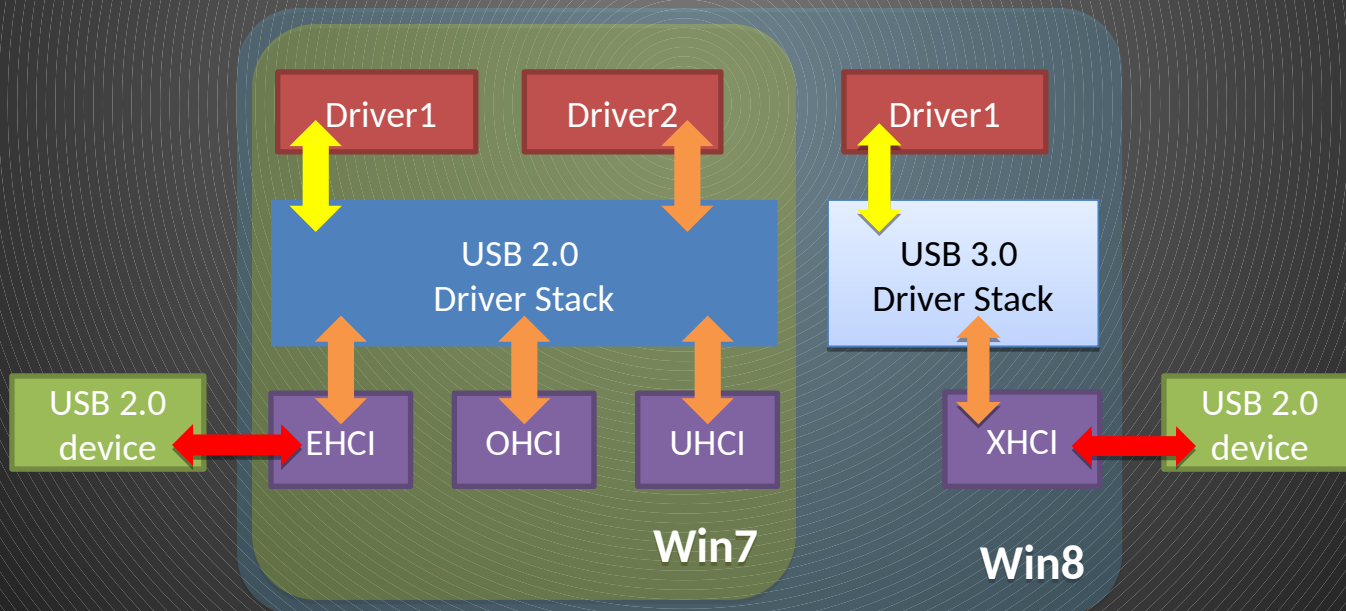
# A Data Transformation Pattern

- Use off-the-shelf techniques to mine patterns
  - Item-set mining
  - Temporal pattern mining
  - *Association rule mining*\*
  - *Graph mining*\*
- Use patterns as features
  - Binary/Categorical features: Presence of patterns
  - Numeric features: Properties of patterns

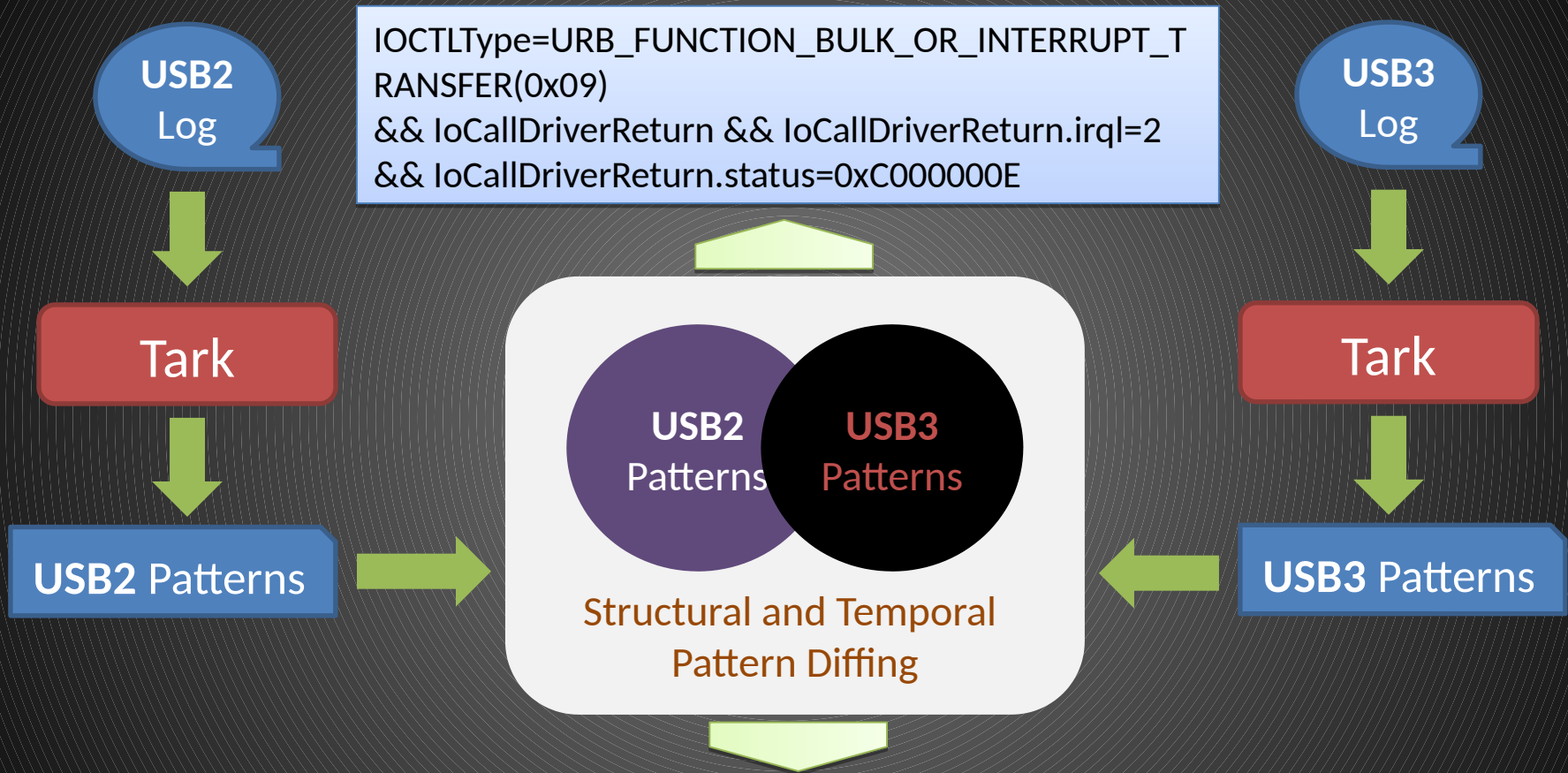
\* We have not tried these pattern mining techniques.

# Example

When a USB 2.0 device is plugged into a USB 3.0 port on Win8, will the USB 3.0 driver in Win8 exhibit the same behavior as the USB 2.0 driver?



# Example



```
IOCTLType=URB_FUNCTION_BULK_OR_INTERRUPT_T  
RANSFER(0x09)  
&& IoCallDriverReturn && IoCallDriverReturn.irql=2  
&& IoCallDriverReturn.status=0xC000000E
```

```
DispatchIrp forward alternates with IrpCompletion && PreIoCompleteRequest  
when  
IOCTLType=IRP_MJ_PNP(0x1B),IRP_MN_START_DEVICE(0x00), irpID=SAME, and  
IrpSubmitDetails.irp.ioStackLocation.control=SAME
```

# Pattern-based Features

## Input Characteristics

- Data is sequential
- Data is structured
- Fields may be irrelevant
- Values may be irrelevant
- Value flow may be relevant

## Output Constraints

- Usable with existing DM/ML algorithms
- Amenable to simple reasoning
- Accessible
- Possess explanatory power

## Pattern

- Use off-the-shelf techniques to mine patterns
  - Item-set mining
  - Temporal pattern mining
- Use patterns as features
  - Binary/Categorical features
  - Numeric features