Windows Memory Dump Analysis

Extended

Extensions, Database and Event Stream Processing,
Data Science and Visualization,
Machine Learning and AI

Version 2.0

Dmitry Vostokov
Software Diagnostics Services

Prerequisites

- Basic WinDbg usage
- Coding in a high-level language
- Ideal previous training:
  - Accelerated Windows Memory Dump Analysis
  - Accelerated .NET Core Memory Dump Analysis
  - Advanced Windows Memory Dump Analysis
  - Accelerated Windows Malware Analysis with Memory Dumps

Why Extended Memory Analysis?

- Limitations of existing commands
- Scripts may be slow or not convenient to use
- Different output format
- Get more insight
- Automate analysis of verbose output

Training Goals

- Review 3rd-party extensions
- Map to memory analysis patterns
- Compare with traditional techniques
- Write our own extensions
- Use data processing, analysis, and visualization
- Use machine learning and AI

Schedule

- Survey of WinDbg extensions
- Writing WinDbg extensions (C, C++, Rust)
- Event stream processing
- Database processing
- Data analysis and visualization
- Machine learning and AI

Training Principles

- Talk only about what I can show
- Lots of pictures
- Lots of examples
- Original content and examples

Course Idea

- Awesome WinDbg Extensions list
- My experience with Kafka
- My experience with JSON data processing
- My interest in data science and visualization (trace and log analysis)
- My interest in machine learning and AI

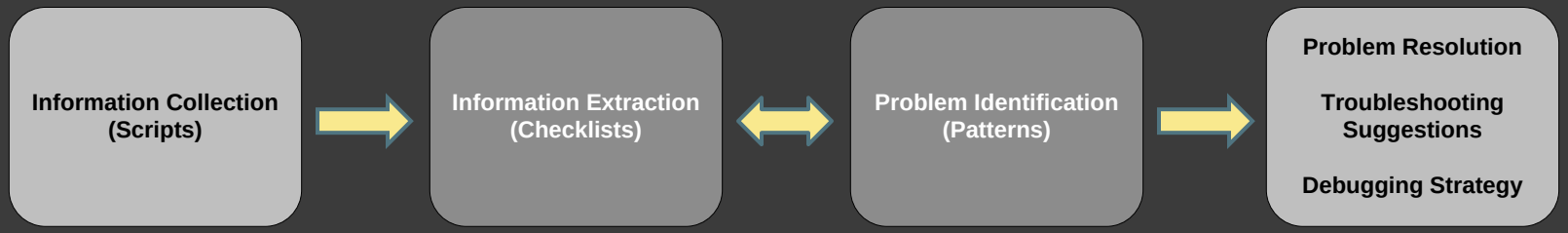
Pattern-Oriented Diagnostic Analysis

Diagnostic Pattern: a common recurrent identifiable problem together with a set of recommendations and possible solutions to apply in a specific context.

Diagnostic Problem: a set of indicators (symptoms, signs) describing a problem.

Diagnostic Analysis Pattern: a common recurrent analysis technique and method of diagnostic pattern identification in a specific context.

Diagnostics Pattern Language: common names of diagnostic and diagnostic analysis patterns. The same language for any operating system: Windows, macOS, Linux, ...



Checklist: http://www.dumpanalysis.org/windows-memory-analysis-checklist

Links

- Applications:

Download links are in the exercise E0.

- Exercise Transcripts:

Included in this book.

Exercise E0

- **Goal:** Install WinDbg or Debugging Tools for Windows, or pull Docker image, and check that symbols are set up correctly
- **Memory Analysis Patterns:** Stack Trace; Incorrect Stack Trace
- \EWMDA\Exercise-E0.pdf

Survey of WinDbg Extensions

Exercises ES1 – ES10

Criteria

- General usefulness for dump analysis
- Addresses common manual techniques
- Corresponds to certain analysis patterns

Exercise ES1

- **Goal:** Explore Patterns WinDbg extension
- \EWMDA\Exercise-ES1.pdf

Exercise ES2

- **Goal:** Explore MEX WinDbg extension
- **Memory Analysis Patterns:** Zombie Processes; Instrumentation Information; Blocked Thread (Software); Active Thread; Suspended Thread; Wait Chain (ALPC); Input Thread; Exception Stack Trace; Stack Trace Collection (Predicate); Stack Trace Collection (CPUs); Spiking Thread; Execution Residue (Unmanaged Space)
- \EWMDA\Exercise-ES2.pdf

Exercise ES3

- **Goal:** Explore DbgKit WinDbg extension
- **Memory Analysis Patterns:** Module Collection; Historical Information; Driver Device Collection; Stack Trace (I/O devices); Stack Trace Collection (I/O drivers); System Object; Value References; Zombie Processes; Virtualized Process (WOW64); Stack Trace Collection; Environment Hint; Deviant Token; Raw Pointer; Out-of-Module Pointer
- \EWMDA\Exercise-ES3.pdf

Exercise ES4

- **Goal:** Explore win32kext WinDbg extension
- **Memory Analysis Patterns:** Handle Limit (GDI, Kernel Space); Wait Chain (Window Messaging)
- \EWMDA\Exercise-ES4.pdf

Exercise ES5

- **Goal:** Explore SwishDbgExt WinDbg extension
- **Memory Analysis Patterns:** Historical Information; Missing Thread (Kernel Space); Driver Device Collection; Patched Code; Out-of-Module Pointer; Self-Diagnosis (Registry); System Object; Namespace
- \EWMDA\Exercise-ES5.pdf

Exercise ES6

- **Goal:** Explore 0cchext WinDbg extension
- **Memory Analysis Patterns:** Execution Residue (Unmanaged Space); Namespace; Context Pointer; Rough Stack Trace (Unmanaged Space); Step Dumps; Evental Dumps
- \EWMDA\Exercise-ES6.pdf

Exercise ES7

- **Goal:** Explore pykd WinDbg extension
- **Memory Analysis Patterns:** Execution Residue (Unmanaged Space)
- \EWMDA\Exercise-ES7.pdf

Raw Stack Analysis

- Symbolic hints at past behavior
- Past stack traces
- Errors, strings, pointers, pointers to pointers

Exercise ES8

- **Goal:** Explore snapshot WinDbg extension
- **Memory Analysis Patterns:** Active Thread
- \EWMDA\Exercise-ES8.pdf

Exercise ES9

- **Goal:** Explore WinDbg Copilot
- **Memory Analysis Patterns:** Analysis Summary; Annotated Stack Trace; Disassembly Summary; Region Summary
- \EWMDA\Exercise-ES9.pdf

Exercise ES10

- **Goal:** Explore ChatDBG WinDbg extension
- **Memory Analysis Patterns:** Frame Trace
- \EWMDA\Exercise-ES10.pdf

Writing WinDbg Extensions

Exercises EW1 – EW4

Goal

- Survey different ways to write extensions
- Simple clean skeletons for further extension
- Useful functionality for analysis patterns

Exercise EW1

- **Goal:** Write WinDbg extension using WdbgExts C API
- \EWMDA\Exercise-EW1.pdf

Exercise EW2

- **Goal:** Write WinDbg extension using DbgEng COM API
- \EWMDA\Exercise-EW2.pdf

Exercise EW3

- **Goal:** Write WinDbg extension using ExtExtension C++ API
- \EWMDA\Exercise-EW3.pdf

Exercise EW4

- **Goal:** Write WinDbg extension using Rust
- \EWMDA\Exercise-EW4.pdf

Event Stream Processing

Exercises EP1 – EP3

Apache Kafka

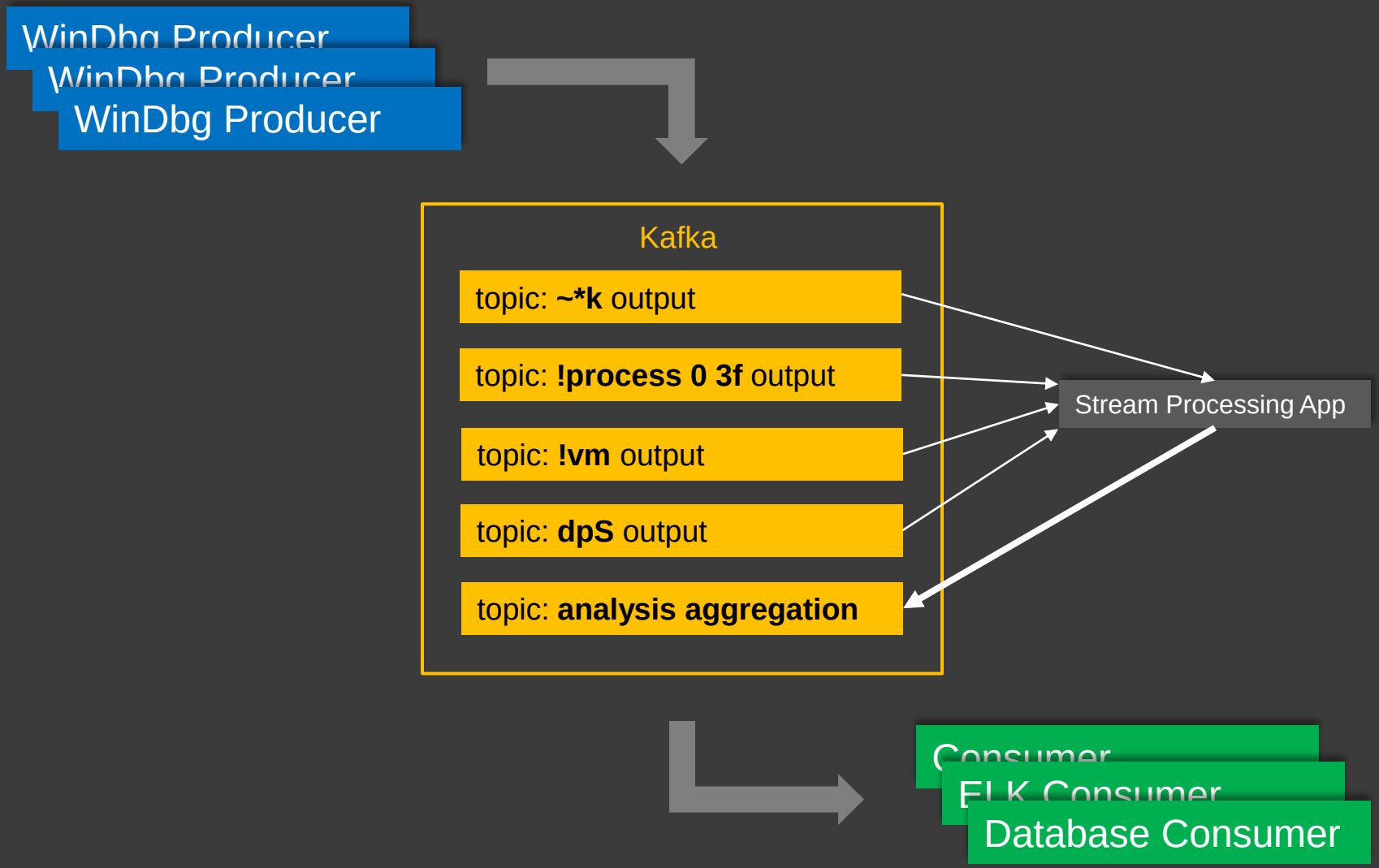
- WinDbg log

- log store

- log processing

- log consumers

Command Logs



Exercise EP1

- **Goal:** Install Apache Kafka and verify that it works correctly
- \EWMDA\Exercise-EP1.pdf

Exercise EP2

- **Goal:** Connect WinDbg to Kafka for logging to various topics
- **Memory Analysis Patterns:** Structure Sheaf; Stack Trace (Command); Stack Trace Collection (Commands)
- \EWMDA\Exercise-EP2.pdf

Exercise EP3

- **Goal:** Configure Kafka Connect to send WinDbg output
- \EWMDA\Exercise-EP3.pdf

Database Processing

Exercises ED1 – ED2

MongoDB

- WinDbg logs as NoSQL data
- Collections of command output, for example, !analyze -v or ~*k
- Command output with added metadata as a document

Exercise ED1

- **Goal:** Install MongoDB and verify that it works correctly
- \EWMDA\Exercise-ED1.pdf

Exercise ED2

- **Goal:** Connect WinDbg to MongoDB for storing analysis documents
- \EWMDA\Exercise-ED2.pdf

Data Science and Visualization

Exercises EV1 – EV3

Pandas

- Tabular raw stack or heap data
- Thousands of rows per thread and millions per heap
- Hundreds of threads

Exercise EV1

- **Goal:** Install Jupyter Notebook and verify that it works correctly
- \EWMDA\Exercise-EV1.pdf

Exercise EV2

- **Goal:** Explore various execution residue statistics and visualization opportunities using Pandas and Matplotlib
- **Memory Analysis Patterns:** Execution Residue (Unmanaged Space); Region Profile; Region Clusters; Namespace
- \EWMDA\Exercise-EV2.pdf

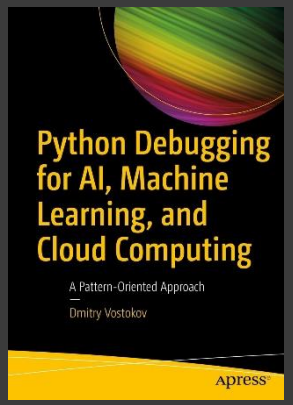
Exercise EV3

- **Goal:** Find anomalies in stack traces from different executions
- **Memory Analysis Patterns:** Rough Stack Trace Collection (Unmanaged Space); Namespace
- \EWMDA\Exercise-EV3.pdf

Diagnostics Presentation Patterns

- Introduced in:

Pattern-Oriented Debugging Process



- Include visualization

- New forthcoming pattern catalog

Machine Learning and AI

Exercises ML1 – ML3

Exercise ML1

- **Goal:** Compute the similarity of stack traces from different executions
- **Memory Analysis Patterns:** Stack Trace Collection (Unmanaged Space); Rough Stack Trace Collection (Unmanaged Space)
- **Feature Extraction:** Function Call Frequencies; N-grams; TF-IDF
- **Comparison:** Cosine Similarity; Jaccard Similarity
- \EWMDA\Exercise-ML1.pdf

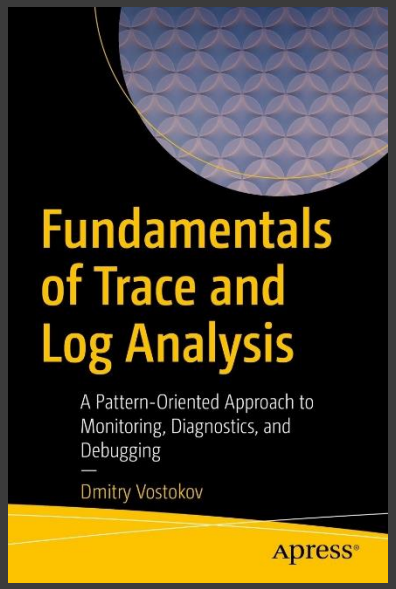
Exercise ML2

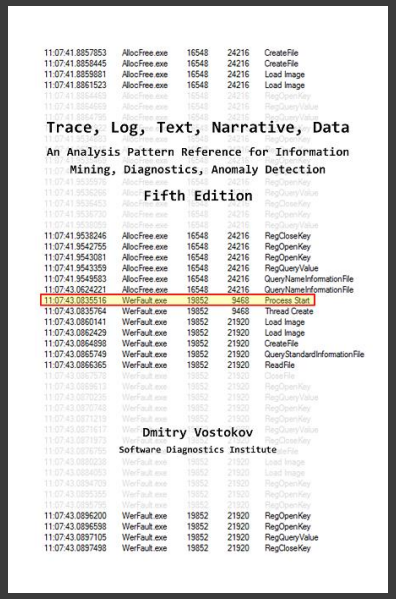
- **Goal:** Group similar stack traces into clusters
- **Memory Analysis Patterns:** Thread Cluster; Stack Trace Collection (Unmanaged Space); Rough Stack Trace Collection (Unmanaged Space)
- **Feature Extraction:** TF-IDF
- **Comparison:** K-means; DBSCAN
- \EWMDA\Exercise-ML2.pdf

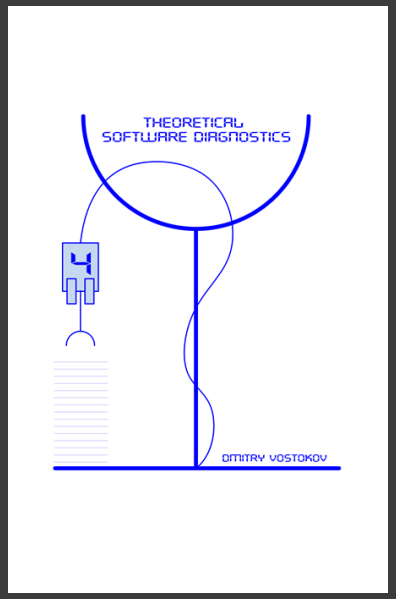
Exercise ML3

- **Goal:** Identify anomalous stack traces
- **Memory Analysis Patterns:** Stack Trace Collection (Unmanaged Space); Rough Stack Trace Collection (Unmanaged Space)
- **Feature Extraction:** TF-IDF
- **Comparison:** Isolation Forest; Autoencoders
- \EWMDA\Exercise-ML3.pdf

Stack Traces as Logs







Further Methods for Stack Traces

- Edit distance
- Sequence alignment
- Word2Vec embeddings
- Recurrent Neural Networks (RNNs)
- Long Short-Term Memory (LSTM) networks
- Transformers
- Graph similarity
- Graph embeddings

Further Course

Machine Learning for Software Diagnostics

Memory Analysis Pattern Links

Execution Residue (Unmanaged Space, User)
Execution Residue (Unmanaged Space, Kernel)
Instrumentation Information
Driver Device Collection
Stack Trace Collection (I/O drivers)
Stack Trace Collection (Predicate)
Stack Trace Collection (CPUs)
Handle Limit (GDI, Kernel Space)
Virtualized Process (WOW64)
Stack Trace Collection (Unmanaged Space)
Missing Thread (Kernel Space)
Wait Chain (Window Messaging)
Self-Diagnosis (Registry)
Out-of-Module Pointer
Deviant Token
Patched Code
Step Dumps
Region Profile
Stack Trace Collection (Commands)
Stack Trace (Command)
Rough Stack Trace (Unmanaged Space)
Annotated Stack Trace
Region Summary
Rough Stack Trace Collection (Unmanaged Space)
Zombie Processes
Module Collection
Historical Information
Stack Trace (I/O devices)
Exception Stack Trace
Blocked Thread (Software)
Active Thread
Suspended Thread
Wait Chain (ALPC)
Spiking Thread
System Object
Input Thread
Value References
Environment Hint
Raw Pointer
Context Pointer
Evental Dumps
Region Clusters
Namespace
Structure Sheaf
Analysis Summary
Disassembly Summary
Frame Trace
Thread Cluster

Resources

- WinDbg Help / WinDbg.org (quick links and some extensions)
- DumpAnalysis.org / SoftwareDiagnostics.Institute / PatternDiagnostics.com
- Software Diagnostics Library
- Comprehensive WinDbg extension collection
- Kafka in Action / Kafka: The Definitive Guide / Apache Kafka
- MongoDB
- Pandas / ydata-profiling / Matplotlib / Polars / Pandas for Everyone / Scikit-learn / TensorFlow
- Machine Learning with Python for Everyone / Modern Deep Learning for Tabular Data
- The Science of Deep Learning / Large Language Models: A Deep Dive
- Text as Data: A New Framework for Machine Learning and the Social Sciences
- Memory Dump Analysis Anthology (Diagnomicon)

Q&A

Please send your feedback using the contact form on PatternDiagnostics.com

Thank you for attendance!