

Ampire 360°



ПЕРСПЕКТИВНЫЙ
МОНИТОРИНГ

Ampire 360°

ПОЛИГОН КИБЕРБЕЗОПАСНОСТИ



Threat Intelligence

SOC & DFIR

Offensive Security

Киберучения

ИИ и киберзащита

Киберполигон Ampire

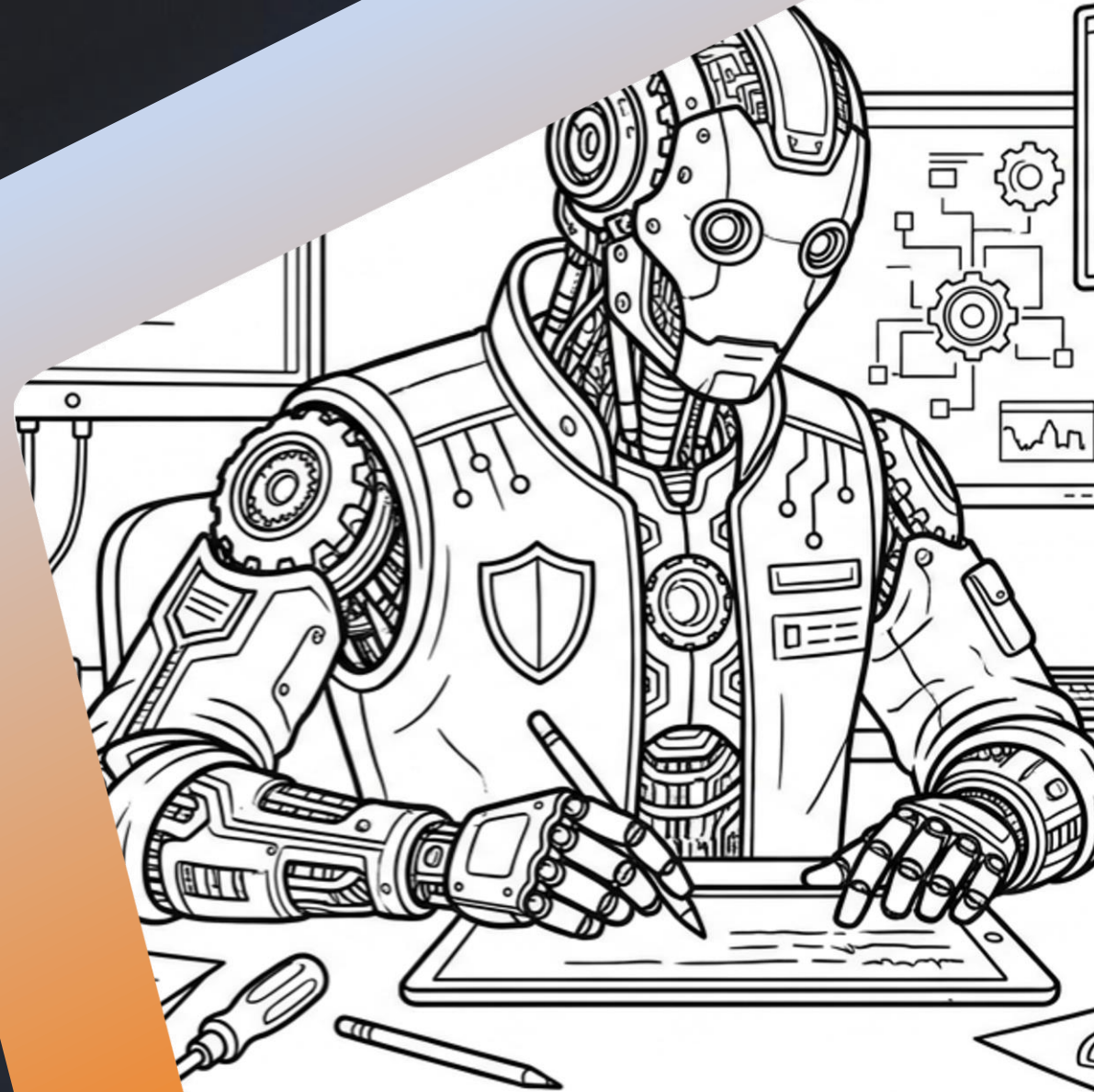
📅 14 октября





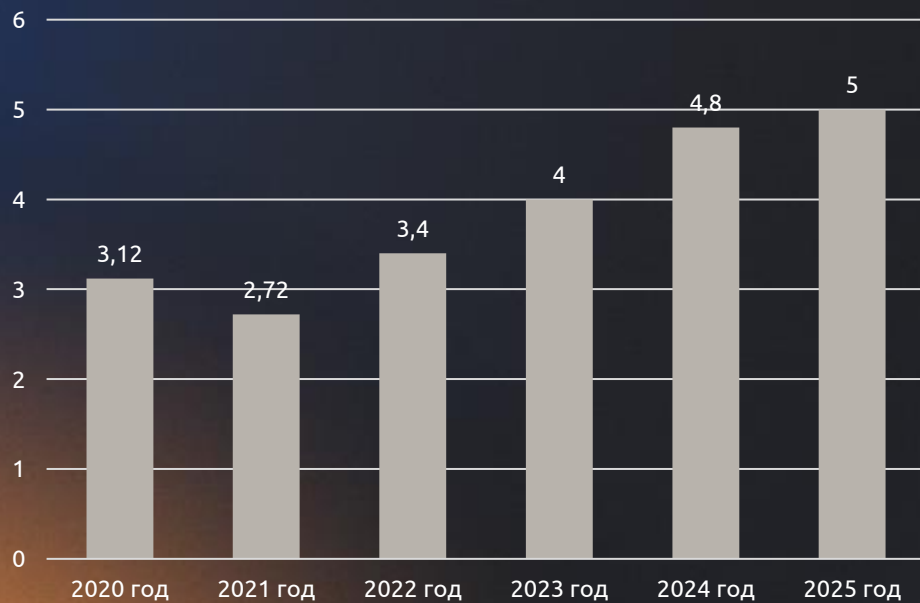
Генерация YARA- правил с помощью LLM

Александр Ташкинов
Старший специалист по исследованию
киберугроз
Перспективный мониторинг



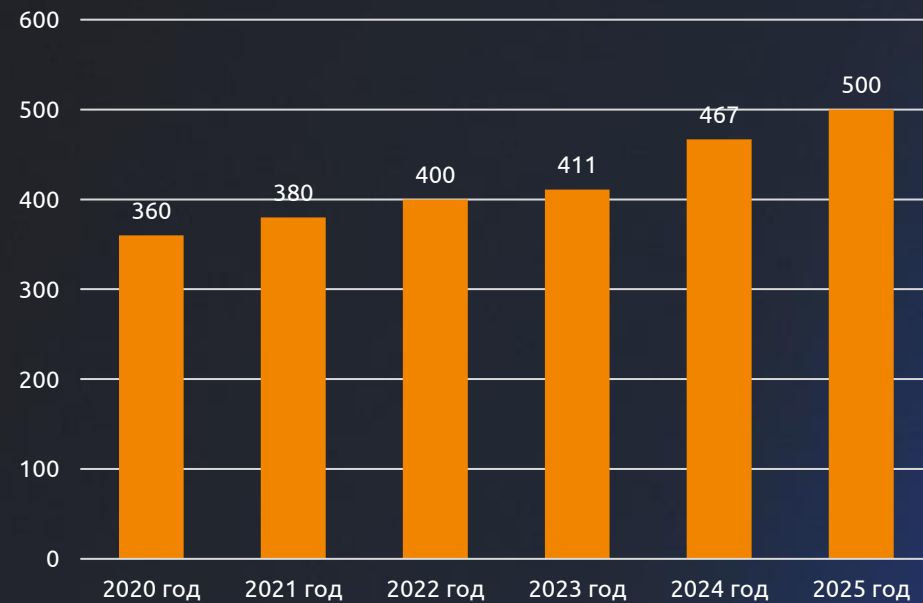
Главные проблемы рынка ИБ

Дефицит кадров в млн чел.



Источник: www.isc2.org

Файлов в сутки в тыс.



Источник: www.kaspersky.ru

Решение проблемы

01

Увеличение штата сотрудников

02

Автоматизация процесса разработки
ЭД

Что такое Yara?

Это инструмент и язык описания паттернов (шаблонов) для поиска в данных

Как найти нужный паттерн в файлах, памяти, процессах или сети?

«Швейцарский нож для исследователей вредоносного ПО»



Общая структура генератора

Физическое
хранилище

Хранилище
вредоносного и
легитимного ПО

БД с записями о
ВПО

MongoDB,
содержащая в себе
записи о ВПО

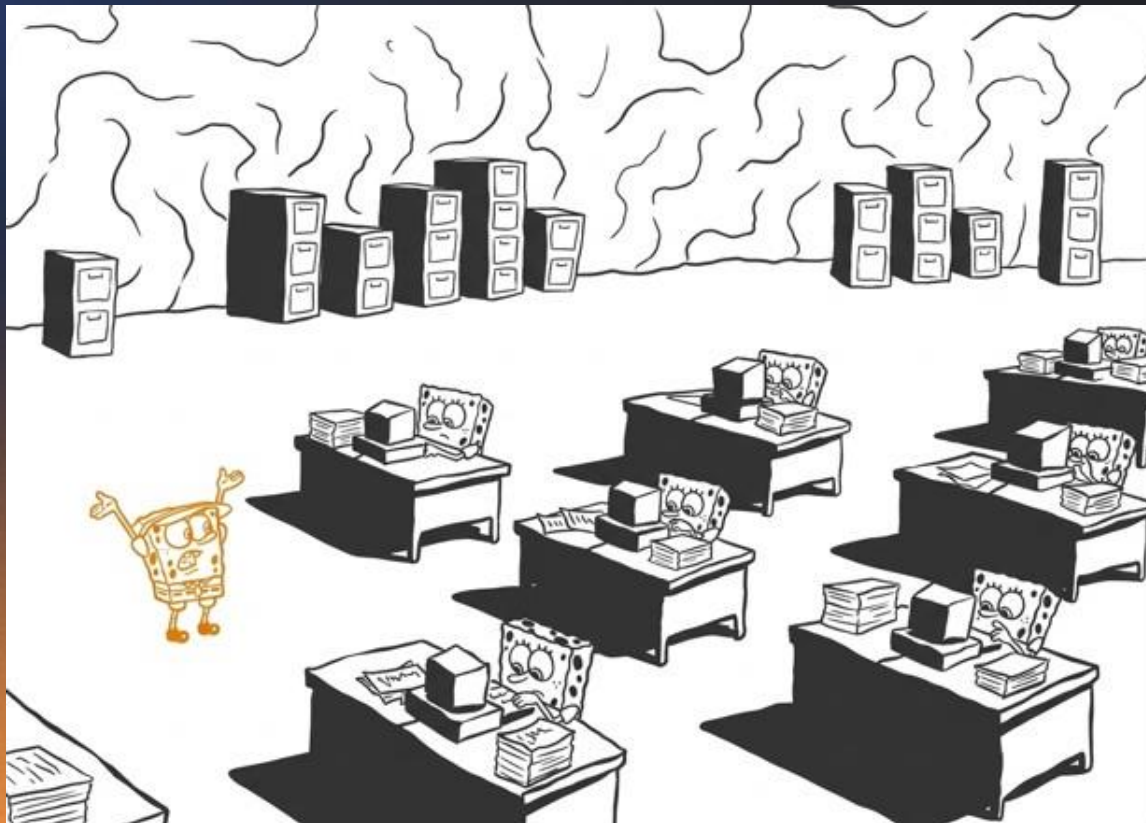
LLM+MCP

Модель и MCP
сервер как мозг
генератора

Кодовая база

База скриптов для
обработки данных
между модулями

Хранилище сэмплов и MongoDB



Как обеспечить
структурный подход
к обрабатываемым
образцам?

Структура MongoDB

Поле _id

```
_id: ObjectId('692970d6f2c51cd96184618a')
file_path : "/samples/Bazaar_Daily/samples_daily/2020-02-24/59fc347dac3dd1c78d62393..."
sha256 : "59fc347dac3dd1c78d62393589818b5417ca041d697d155040988b14562bc797"
size_bytes : 120320
extension : "exe"
yara_raw : "None"
tested : "False"
affected_os : "Windows"
```

Обязательный
уникальный
идентификатор каждого
документа

Структура MongoDB

Поле file_path

```
_id: ObjectId('692970d6f2c51cd96184618a')  
file_path: "/samples/Bazaar_Daily/samples_daily/2020-02-24/59fc347dac3dd1c78d62393..."  
sha256: "59fc347dac3dd1c78d62393589818b5417ca041d697d155040988b14562bc797"  
size_bytes: 120320  
extension: "exe"  
yara_raw: "None"  
tested: "False"  
affected_os: "Windows"
```

Абсолютный путь к
файлу в хранилище

Структура MongoDB

Поле sha256

```
_id: ObjectId('692970d6f2c51cd96184618a')  
file_path : "/samples/Bazaar_Daily/samples_daily/2020-02-24/59fc347dac3dd1c78d62393..."  
sha256 : "59fc347dac3dd1c78d62393589818b5417ca041d697d155040988b14562bc797"  
size_bytes : 120320  
extension : "exe"  
yara_raw : "None"  
tested : "False"  
affected_os : "Windows"
```

➤ Хэш в формате SHA256

Структура MongoDB

Поле size_bytes

```
_id: ObjectId('692970d6f2c51cd96184618a')
file_path : "/samples/Bazaar_Daily/samples_daily/2020-02-24/59fc347dac3dd1c78d62393..."
sha256 : "59fc347dac3dd1c78d62393589818b5417ca041d697d155040988b14562bc797"
size_bytes : 120320
extension : "exe"
yara_raw : "None"
tested : "False"
affected_os : "Windows"
```

Размер сэмпла в
байтах

Структура MongoDB

Поле extension

```
_id: ObjectId('692970d6f2c51cd96184618a')
file_path : "/samples/Bazaar_Daily/samples_daily/2020-02-24/59fc347dac3dd1c78d62393..."
sha256 : "59fc347dac3dd1c78d62393589818b5417ca041d697d155040988b14562bc797"
size_bytes : 120320
extension : "exe"
yara_raw : "None"
tested : "False"
affected_os : "Windows"
```

Расширение
файла

Структура MongoDB

Поле yara_raw

```
_id: ObjectId('692970d6f2c51cd96184618a')
file_path : "/samples/Bazaar_Daily/samples_daily/2020-02-24/59fc347dac3dd1c78d62393..."
sha256 : "59fc347dac3dd1c78d62393589818b5417ca041d697d155040988b14562bc797"
size_bytes : 120320
extension : "exe"
yara_raw : "None"
tested : "False"
affected_os : "Windows"
```

Готовое уага
правило

Структура MongoDB

Поле tested

```
_id: ObjectId('692970d6f2c51cd96184618a')
file_path : "/samples/Bazaar_Daily/samples_daily/2020-02-24/59fc347dac3dd1c78d62393..."
sha256 : "59fc347dac3dd1c78d62393589818b5417ca041d697d155040988b14562bc797"
size_bytes : 120320
extension : "exe"
yara_raw : "None"
tested : "False"
affected_os : "Windows"
```

Результат
тестирования

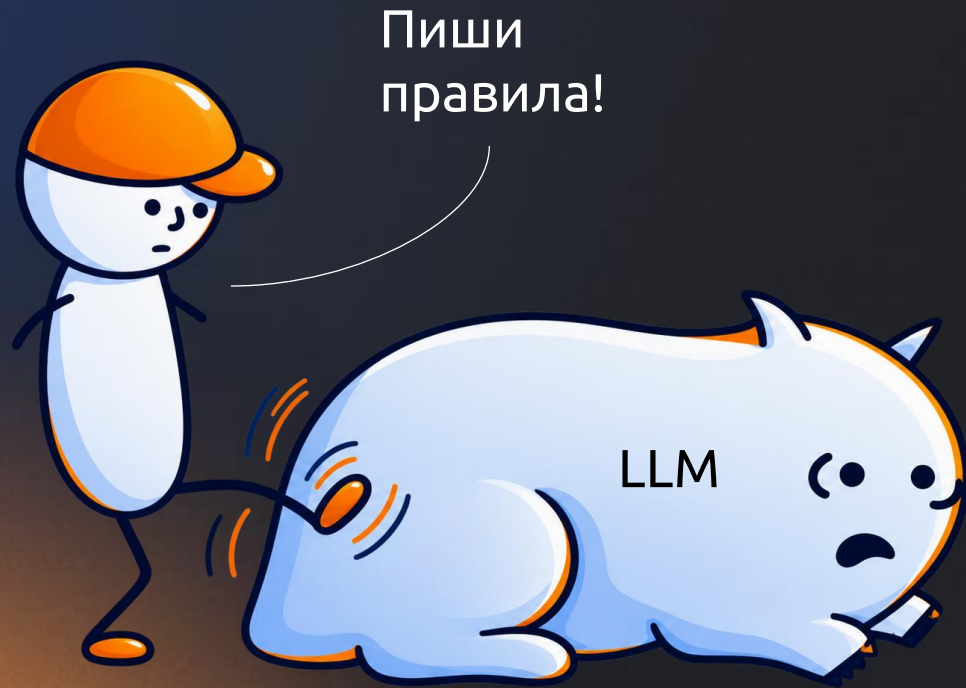
Структура MongoDB

Поле affected_os

```
_id: ObjectId('692970d6f2c51cd96184618a')  
file_path : "/samples/Bazaar_Daily/samples_daily/2020-02-24/59fc347dac3dd1c78d62393..."  
sha256 : "59fc347dac3dd1c78d62393589818b5417ca041d697d155040988b14562bc797"  
size_bytes : 120320  
extension : "exe"  
yara_raw : "None"  
tested : "False"  
affected_os : "Windows"
```

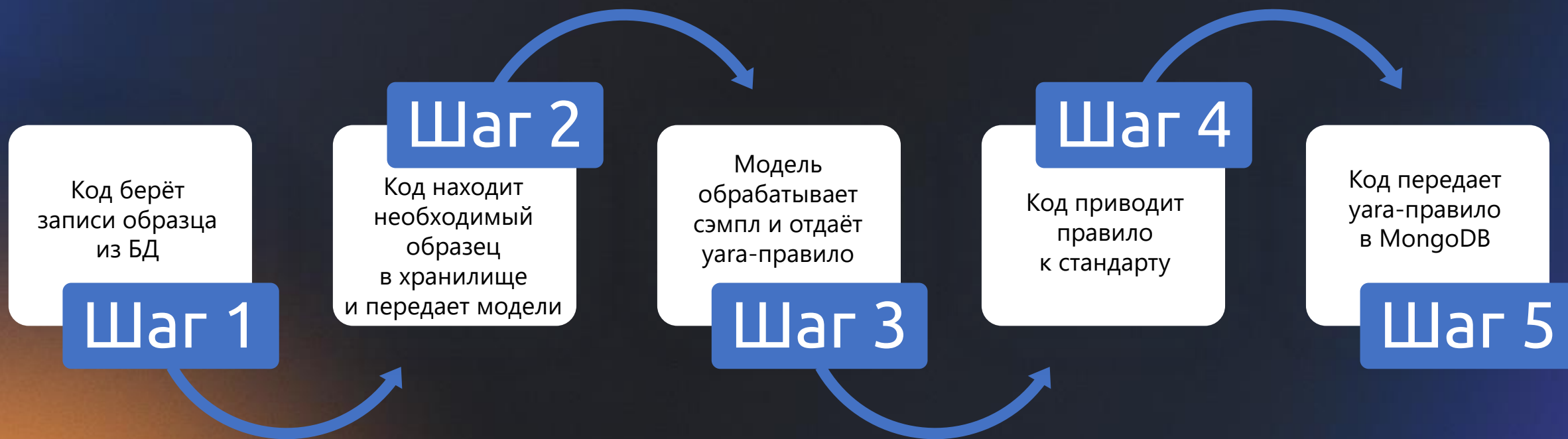
Затронутая ОС

Модель и вспомогательные скрипты



Мозг всего
генератора

Как происходит генерация правила



Результат генерации

```
_id: ObjectId('6a8c20b3b4a227890dbb0a0a'),
file_path: '/samples/Bazaar_Daily/samples_daily/2026-08-01/7
sha256: '781965934626c16eea2cb5ffe657bb52ffa23b001af8c3b26a1
size_bytes: 5632354,
extension: 'exe',
yara_raw: 'rule exe_auto_781965934626c16eea2cb5ffe657bb52ffa
'\n' +
'  meta:\n' +
'    description = "Auto: malicious exe file"\n' +
'    reference = "bazaar.abuse.ch/sample/781965934626c
'    techniques = "T1204.002,T1059.006,T1027,T1105"\n'
'    capec = ""\n' +
'    cwe = ""\n' +
'    cve = ""\n' +
'  strings:\n' +
'    $s1 = "cryptor" ascii\n' +
'    $s2 = "pyiboot01_bootstrap" ascii\n' +
'    $s3 = "mpyimod04_pywin32" ascii\n' +
'    $s4 = "python39.dll" ascii\n' +
'    $s5 = "libcrypto-1_1.dll" ascii\n' +
'    $s6 = "_MEIPASS" ascii\n' +
'  condition:\n' +
'    uint16(0) == 0x5A4D and\n' +
'    filesize < 6195590 and\n' +
'    all of them\n' +
'  },
tested: 'True',
affected_os: 'Windows',
is_duplicate: true
```

```
yara_raw: 'rule exe_auto_781965934626c16eea2cb5ffe657bb52ffa23b001af8c3b26a14cf638af50c9a\n' +
'\n' +
'  meta:\n' +
'    description = "Auto: malicious exe file"\n' +
'    reference = "bazaar.abuse.ch/sample/781965934626c16eea2cb5ffe657bb52ffa23b001af8c3b26a14cf638af50c9a"\n' +
'    techniques = "T1204.002,T1059.006,T1027,T1105"\n' +
'    capec = ""\n' +
'    cwe = ""\n' +
'    cve = ""\n' +
'  strings:\n' +
'    $s1 = "cryptor" ascii\n' +
'    $s2 = "pyiboot01_bootstrap" ascii\n' +
'    $s3 = "mpyimod04_pywin32" ascii\n' +
'    $s4 = "python39.dll" ascii\n' +
'    $s5 = "libcrypto-1_1.dll" ascii\n' +
'    $s6 = "_MEIPASS" ascii\n' +
'  condition:\n' +
'    uint16(0) == 0x5A4D and\n' +
'    filesize < 6195590 and\n' +
'    all of them\n' +
'  },
```

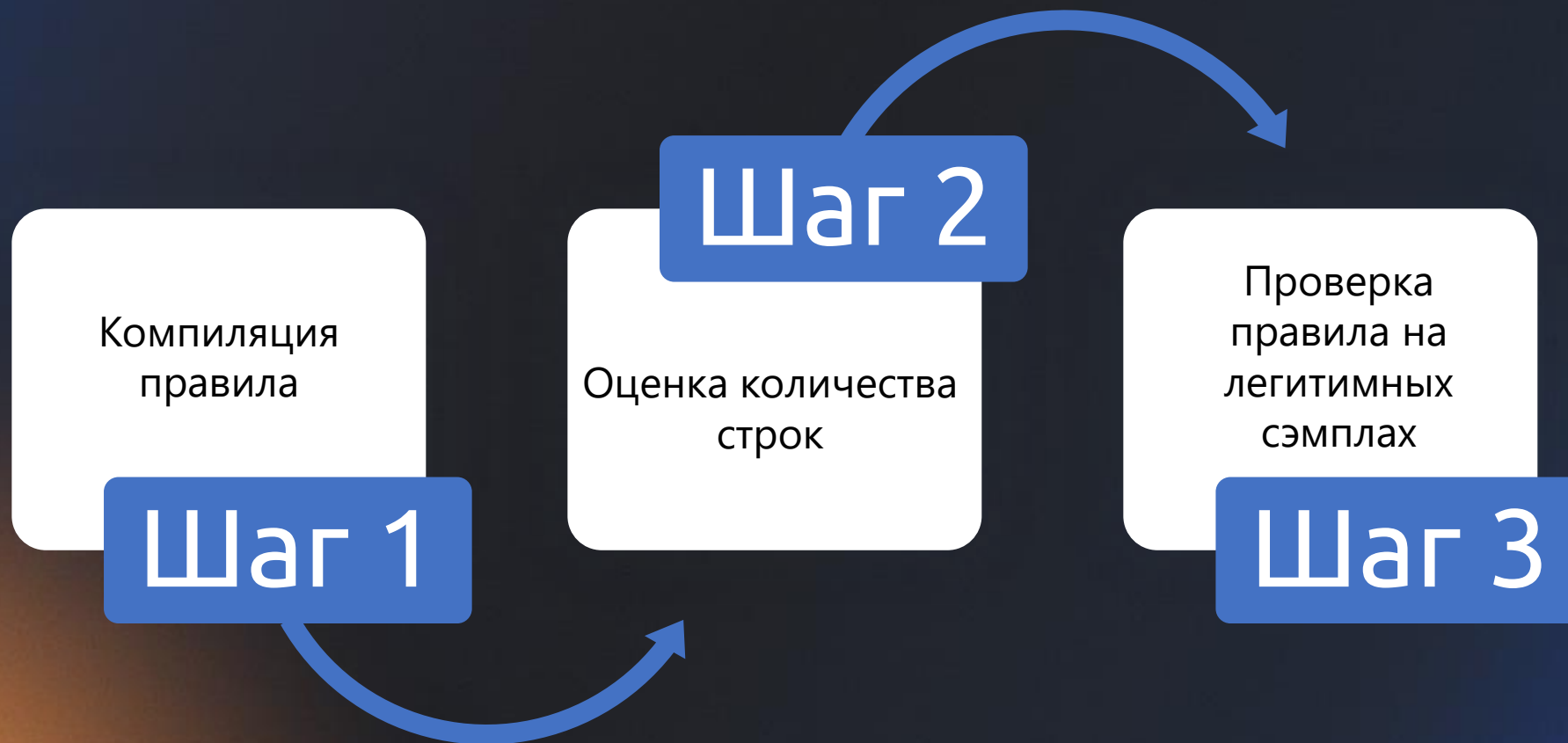


Тестирование



Решаем проблему
синтаксических
ошибок и ложных
срабатываний

Как происходит тестирование правила



Дедупликация



Дубликаты
уменьшают
производительность
и не увеличивают
покрытие

Пять этапов дедупликации

Первый этап

Извлечение секции strings

Второй этап

Нормализация правила

Третий этап

Разбиение на шинглы

Четвёртый этап

Локально-чувствительное хеширование (LSH)

Пятый этап

MinHash — математическая модель

Первый этап

Извлечение секции strings

```
rule dll_auto_9c2838e120c7ed5b582bedc6177f14a52aa578adeea269d0f96fc71a95bd6e68
{
  meta:
    description = "Auto: malicious dll file"
    reference = "bazaar.abuse.ch/sample/9c2838e120c7ed5b582bedc6177f14a52aa578adeea269d0f96fc71a95bd6e68"
    techniques = "T1204.002"
    capec = ""
    cwe = ""
    cve = ""
  strings:
    $s1 = "powershell -windowstyle hidden -command" nocase
    $s2 = "iwr -useb http://147.45.44.131/infopage/iubn.ps1" nocase
    $s3 = "X-Special-Header" nocase
    $s4 = "qInx8F3tuJDHXg0EfPJjbaipYaSE1mobJ2YRyo2rjNgnVDhJvevN8R2ku8oPCBonhmpzFb2GYqPiLhJq"
    $s5 = "DllRegisterServer"
    $s6 = "!This program cannot be run in DOS mode."
    $s9 = "147.45.44.131"
    $s10 = "iex" nocase
  condition:
    uint16(0) == 0x5A4D and
    filesize < 99124 and
    all of them
}
```

Было

```
$s1 = "powershell -windowstyle hidden -command" nocase  
$s2 = "iwr -useb http://147.45.44.131/infopage/iubn.ps1" nocase  
$s3 = "X-Special-Header" nocase  
$s4 = "qInx8F3tuJDHXgOEfPJjbaipYaSE1mobJ2YRyo2rjNgnVDhJvevN8R2ku8oPCBonhmpzFb2GYqPiLhJq"  
$s5 = "DllRegisterServer"  
$s6 = "!This program cannot be run in DOS mode."  
$s9 = "147.45.44.131"  
$s10 = "iex" nocase
```



Стало

```
"powershell -windowstyle hidden -command"  
"iwr -useb http://147.45.44.131/infopage/iubn.ps1"  
"X-Special-Header"  
"qInx8F3tuJDHXgOEfPJjbaipYaSE1mobJ2YRyo2rjNgnVDhJvevN8R2ku8oPCBonhmpzFb2GYqPiLhJq"  
"DllRegisterServer"  
"!This program cannot be run in DOS mode."  
"147.45.44.131"  
"iex"
```

"powershell -windowstyle hidden -
command"

Шингл 1

"X-Special-
Header"

Шингл 3

"iwr -useb
<http://147.45.44.131/infopage/iubn.ps1>"

Шингл 2

- Шингл (от англ. shingle — «чешуйка», «гонт») — это фрагмент текста, представляющий собой последовательность слов или субсловных единиц (например, морфем), идущих подряд

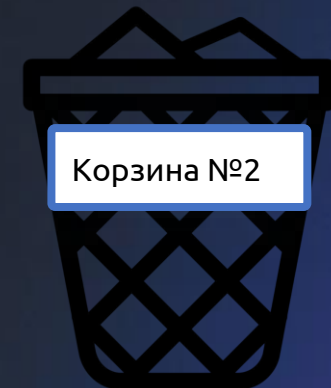
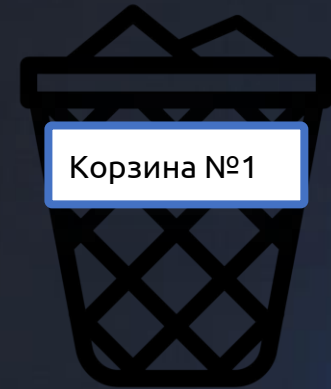
Четвёртый этап

Локально-чувствительное хеширование (LSH)

Rule A = {"MALWARE", "encrypt", "bitcoin", "cmd.exe",
"kernel32.dll"}
= {10, 5, 20, 8, 12}
→ min = 5

Rule B = {"MALWARE", "encrypt", "ransom", "cmd.exe",
"kernel32.dll"}
= {10, 5, 15, 8, 12}
→ min = 5

Rule C = {"Trojan", "download", "http",
"ws2_32.dll"}
= {25, 30, 35, 40}
→ min = 25



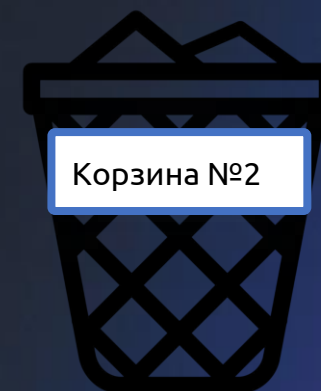
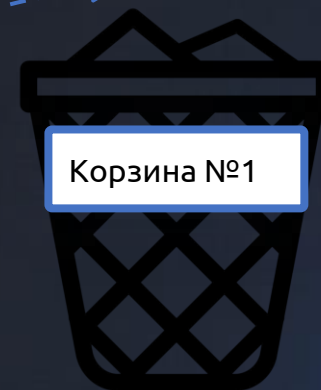
Четвёртый этап

Локально-чувствительное хеширование (LSH)

Rule A = {"MALWARE", "encrypt", "bitcoin", "cmd.exe",
"kernel32.dll"}
= {10, 5, 20, 8, 12}
→ min = 5

Rule B = {"MALWARE", "encrypt", "ransom", "cmd.exe",
"kernel32.dll"}
= {10, 5, 15, 8, 12}
→ min = 5

Rule C = {"Trojan", "download", "http",
"ws2_32.dll"}
= {25, 30, 35, 40}
→ min = 25



Четвёртый этап

Локально-чувствительное хеширование (LSH)

Rule A = {"MALWARE", "encrypt", "bitcoin", "cmd.exe",
"kernel32.dll"}
= {10, 5, 20, 8, 12}
→ min = 5

Rule B = {"MALWARE", "encrypt", "ransom", "cmd.exe",
"kernel32.dll"}
= {10, 5, 15, 8, 12}
→ min = 5

Rule C = {"Trojan", "download", "http",
"ws2_32.dll"}
= {25, 30, 35, 40}
→ min = 25

Корзина №1

Корзина №2

Вероятностный алгоритм, который позволяет быстро оценить, насколько похожи два множества. Его главная идея в том, чтобы не сравнивать все элементы напрямую, а вместо этого построить для каждого множества компактную «сигнатуру» (набор чисел) и уже по ней судить о сходстве.

$$\bullet J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

Коэффициент равен нулю, когда множества не имеют общих элементов, и единице, когда множества равны, в остальных случаях значение где-то посередине.

Пример работы

#63805

Правило А: exe_auto_72e31f7ae295b0cdedfaf6d6db6063485ff39cd0871769d6922b135b1ec8cee7 (id=6a8c20b4b4a227890dbb0a2f, strings=6)

Правило В: exe_auto_5522271f10b6fb5cf4fb54fe7fced441ea3f6a4ace1224e121eda6267dc663cd (id=6a8c20c5b4a227890dbb0c12, strings=5)

Jaccard: 57.14%

Containment: 80.00%

Правило А

```
strings:\n' +  
    $s1 = "Could not load PyInstaller's embedded PKG archive from the executable (%s)"\n' +  
    $s2 = "mpyimod04_pywin32"\n' +  
    $s3 = "Failed to get _MEIPASS as PyObject." \n' +  
    $s4 = "PYINSTALLER_STRICT_UNPACK_MODE"\n' +  
    $s5 = "urllib.request)" ascii\n' +  
    $s6 = "subprocess)" ascii\n' +
```

Правило В

```
strings:\n' +  
    $s1 = "Could not load PyInstaller's embedded PKG archive from the executable (%s)" ascii\n' +  
    $s2 = "mpyimod04_pywin32" ascii\n' +  
    $s3 = "mpyimod03_ctypes" ascii\n' +  
    $s4 = "PYINSTALLER_STRICT_UNPACK_MODE" ascii\n' +  
    $s5 = "Failed to get _MEIPASS as PyObject." ascii\n' +
```

Пример работы

#63805

Правило А: exe_auto_72e31f7ae295b0cdedfaf6d6db6063485ff39cd0871769d6922b135b1ec8cee7 (id=6a8c20b4b4a227890dbb0a2f, strings=6)

Правило В: exe_auto_5522271f10b6fb5cf4fb54fe7fced441ea3f6a4ace1224e121eda6267dc663cd (id=6a8c20c5b4a227890dbb0c12, strings=5)

Jaccard: 57.14%

Containment: 80.00%

Правило А

strings:\n' +

\$s1 = "Could not load PyInstaller's embedded PKG archive from the executable (%s)" \n' +

\$s2 = "mpyimod04_pywin32" \n' +

\$s3 = "Failed to get _MEIPASS as PyObject." \n' +

\$s4 = "PYINSTALLER_STRICT_UNPACK_MODE" \n' +

\$s5 = "urllib.request)" ascii\n' +

\$s6 = "subprocess)" ascii\n' +

Правило В

strings:\n' +

\$s1 = "Could not load PyInstaller's embedded PKG archive from the executable (%s)" ascii\n' +

\$s2 = "mpyimod04_pywin32" ascii\n' +

\$s3 = "mpyimod03_ctypes" ascii\n' +

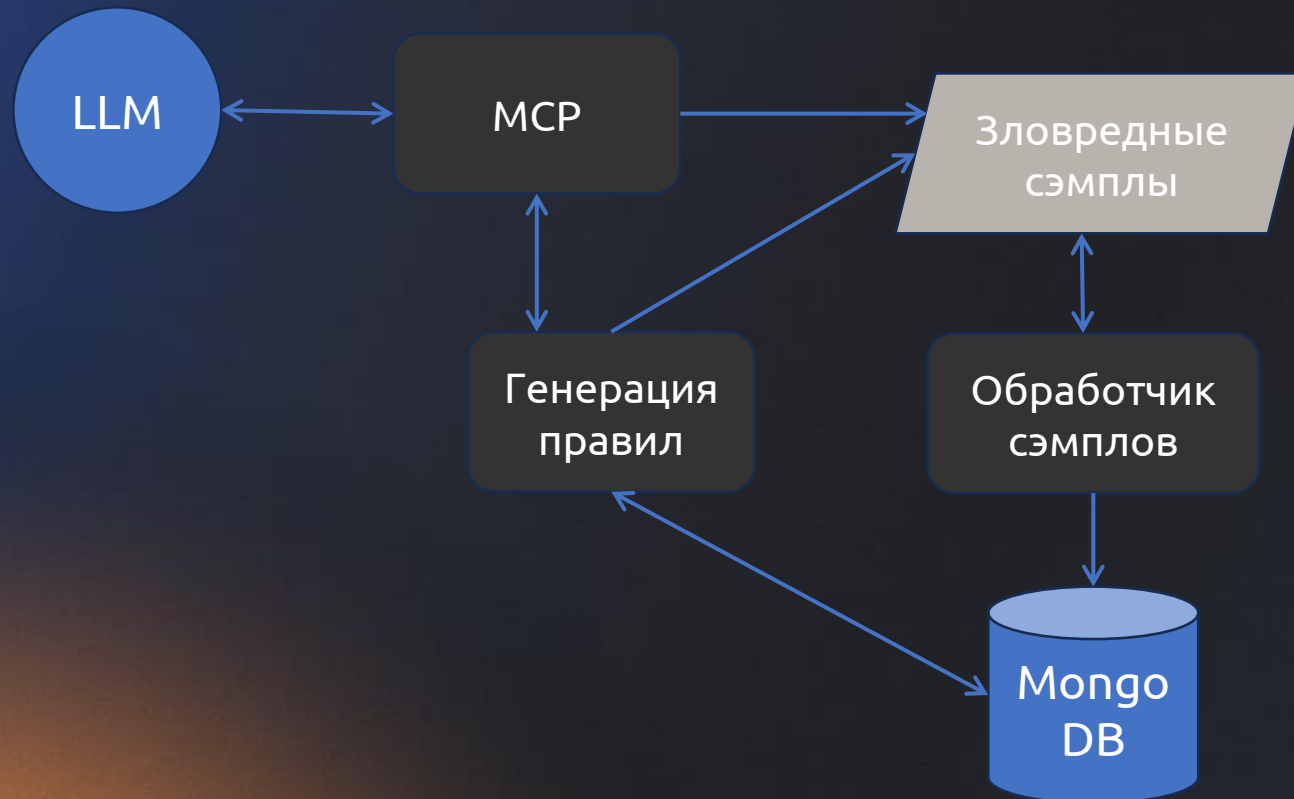
\$s4 = "PYINSTALLER_STRICT_UNPACK_MODE" ascii\n' +

\$s5 = "Failed to get _MEIPASS as PyObject." ascii\n' +

Полный процесс создания набора правил



Полный процесс создания набора правил



Полный процесс создания набора правил



Полный процесс создания набора правил



Проблемы генератора



Потенциальные
сложности
и как их можно
решить

Качество правил

```
yara_raw: 'rule elf_auto_70ffbfd42ba067fb1c8671192da
{
  meta:
    {
      description = "Auto: malicious elf file"
      reference = "bazaar.abuse.ch/sample/70ffbd
      techniques = "T1204.002"
      capec = ""
      cwe = ""
      cve = ""

      strings:
        {
          $upx_info = "This file is packed with the UPX executable packer"
          $upx_id = "UPX 3.94 Copyright (C) 1996-2017 the UPX Team"
          $prot_exec = "PROT_EXEC|PROT_WRITE failed"
          $upx_sf = "http://upx.sf.net"
          $tcp_string = "tcpd"

          condition:
            uint32(0) == 0x464C457F and
            filesize < 35654 and
            all of them
        }
    }
}
```

```
strings:\n' +
```

```
    $upx_info = "This file is packed with the UPX executable packer"\n' +
```

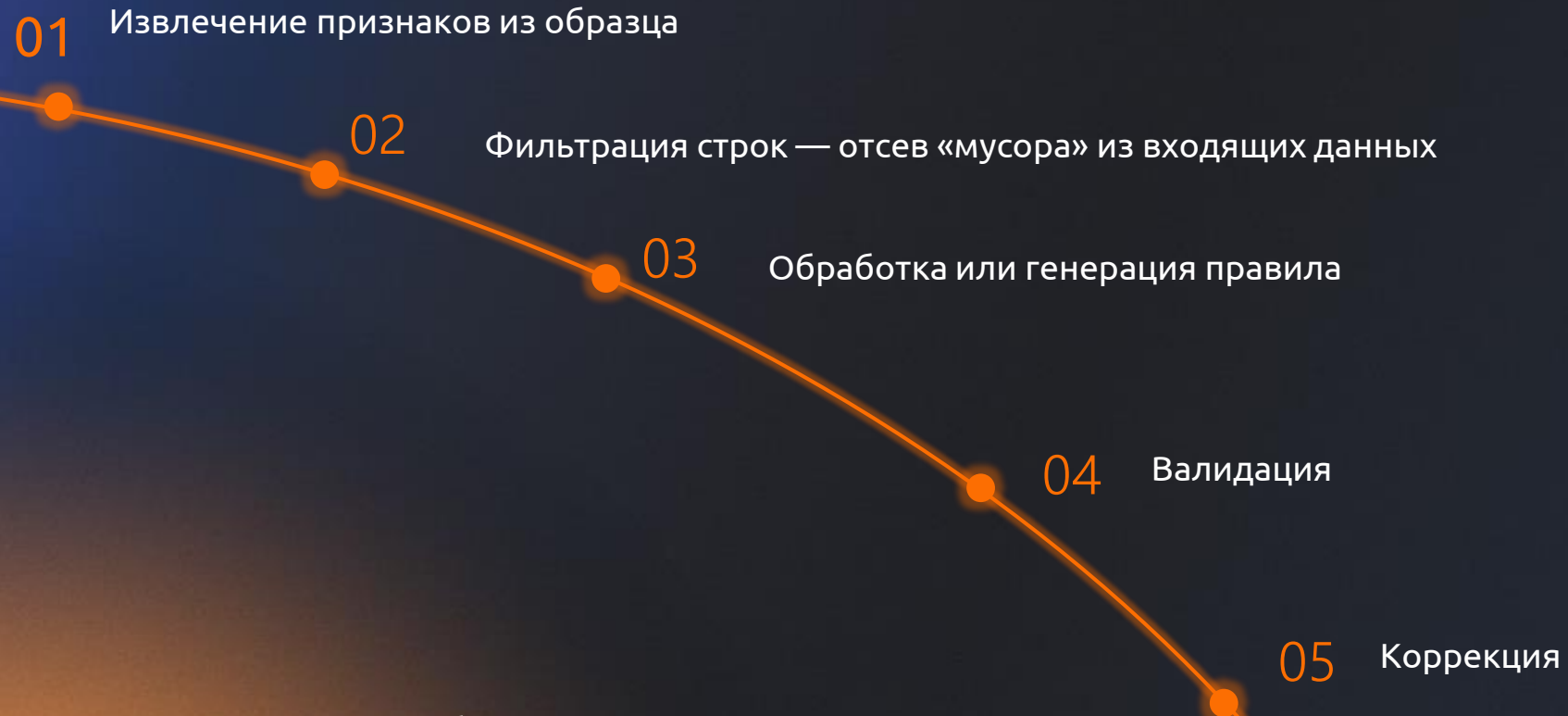
```
    $upx_id = "UPX 3.94 Copyright (C) 1996-2017 the UPX Team"\n' +
```

```
    $prot_exec = "PROT_EXEC|PROT_WRITE failed"\n' +
```

```
    $upx_sf = "http://upx.sf.net"\n' +
```

```
    $tcp_string = "tcpd"\n' +
```

Время работы модели



Под временем работы модели считаем не обработку одного сэмпла, а цепочку процессов

Проблемные сэмплы

Упакованные или
обфусцированные

Ломается извлечение признаков

Очень маленькие или
большие образцы

Мало строк или много «мусора» в правиле

Установщики/Дропперы

Высокий риск ложных срабатываний

Документы (*.docx, *.pdf)

На выходе правила очень узкие и легко обходятся минимальными изменениями

Raw shellcode/без
формата

Модель не может опереться на структурные признаки, путается

Проблемные сэмплы

Скрипты
(VBA/PS1/JS/HTML)

Правила на выходе слишком узкие

Сэмплы с одноразовыми
строками

Правило быстро теряет актуальность и засоряет сборку

Совпадение с
легитимным ПО

Высокий шанс FP

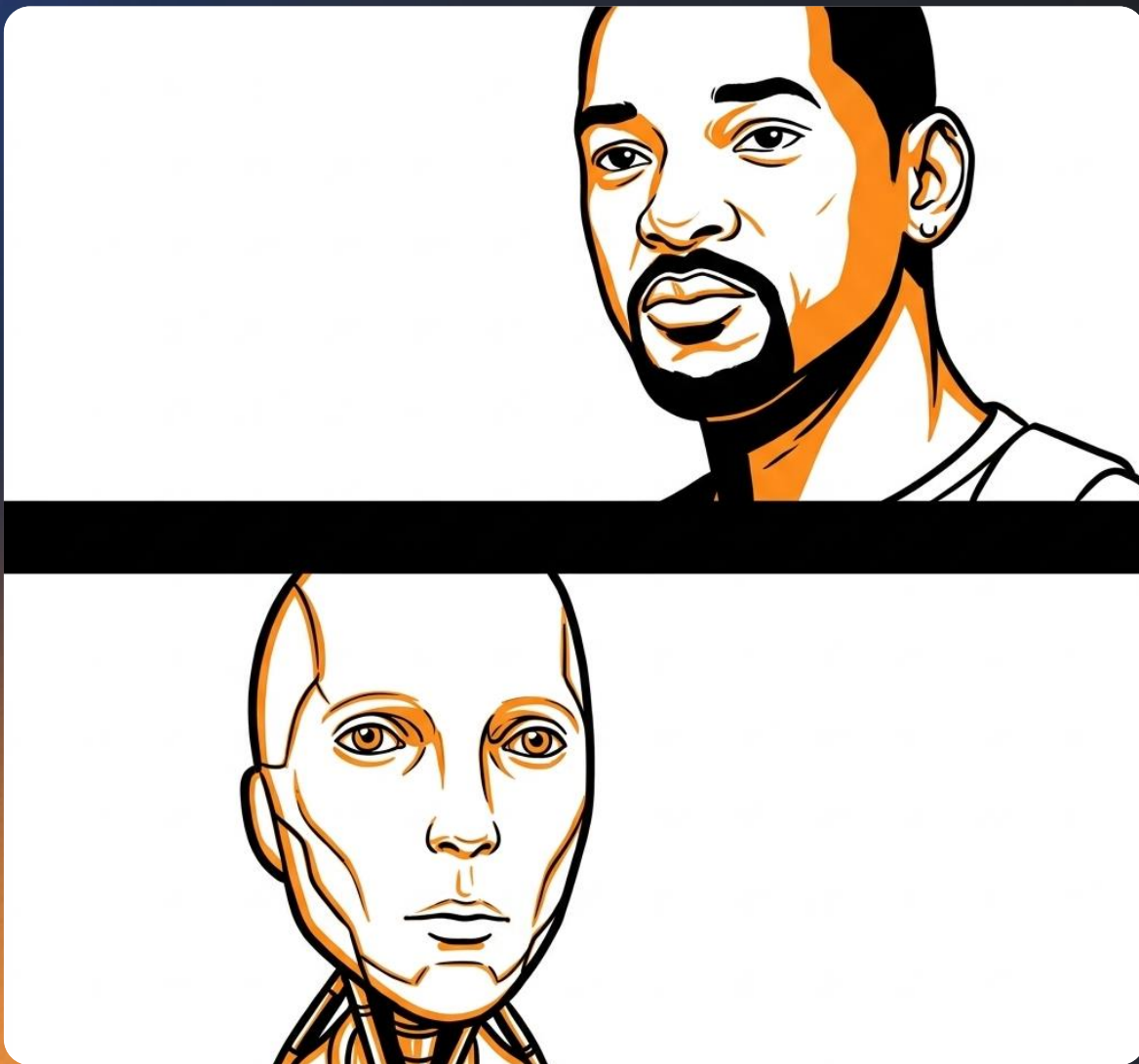
Некорректная разметка

Правила на легитимные образцы

Семейные вариации

Большое количество дубликатов

Применимость «Искусственных» ЭД



Генератор
сочинит
экспертные
данные?
Превратит кусок
строк в уага-
правило?

Сравнение эффективности

Исполнитель	Правил в час	Качество правил	Процент правил в релизе	Процент покрытия образцов
Младший специалист	0.5	7.5	90%	60%
Старший специалист	1.5	10	99%	95%
Генератор	120	6	10–20%	70%

Благодарю за внимание

Александр Ташкинов

Старший специалист по исследованию киберугроз
Перспективный мониторинг



Ampire 360°