Гибридный метод обнаружения шеллкодов

Гайворонская С.А., аспирант 2 г.о.

Научный руководитель:

к.ф.-м.н Гамаюнов Д.Ю.

Ботнеты

Ботнет – множество зараженных узлов (ботов), выполняющих команды управляющего узла (ботмастера).

Вредоносная активность ботнета:

- Организация DDOS атак
- Рассылка спама
- Кража пользовательской информации
- Хостинг ВПО
- …

Способы распространения ботнетов

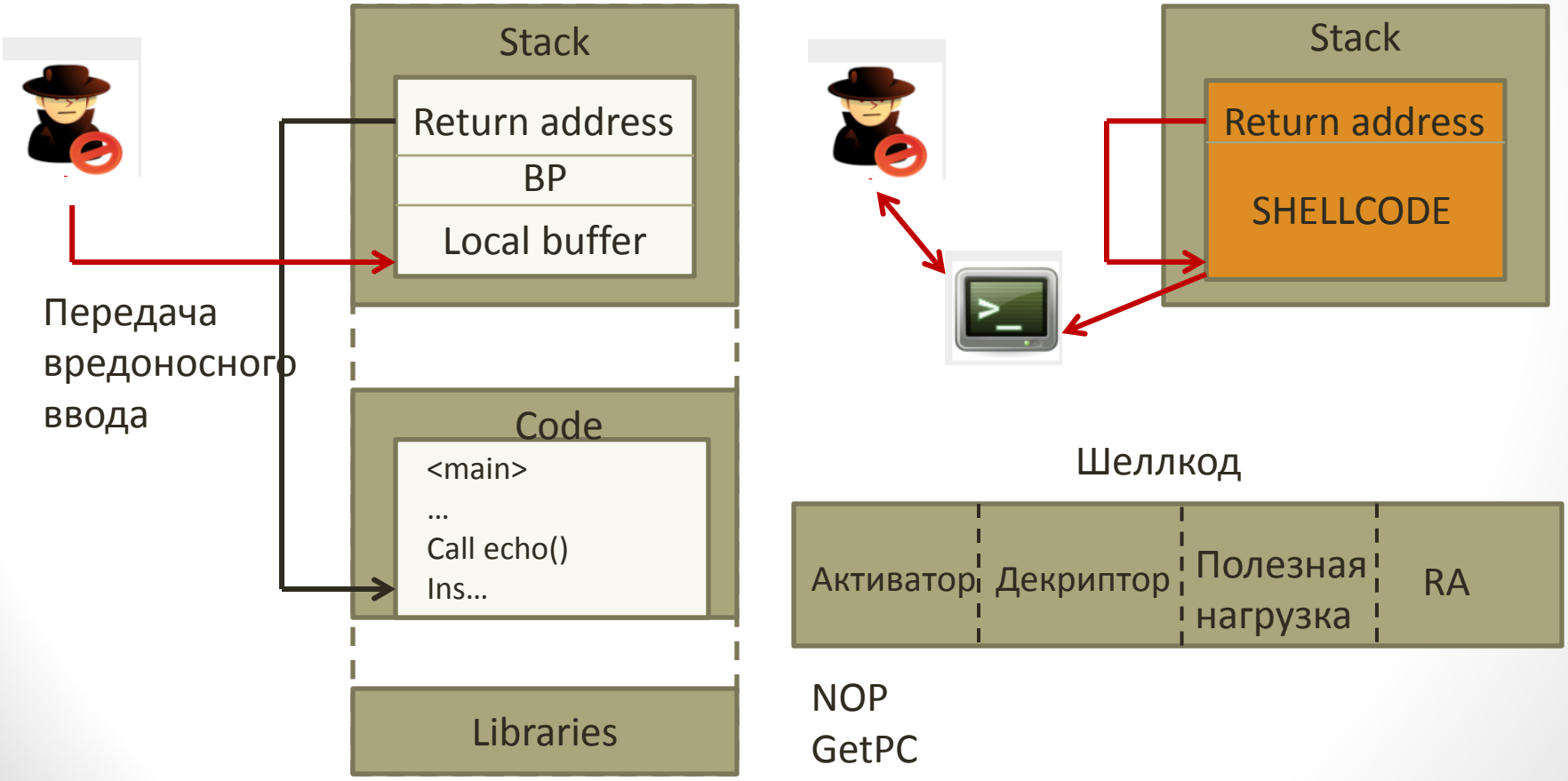
- Remote root – экспулатация удаленных уязвимостей в распростаненном ПО с целью захвата машины-жертвы
- Drive-by-download – экспуатация уязвимостей в браузерах
- Действия пользователя

Направление исследования

- Обнаружение феномена распространия ботнетов посредством сетевых червей
- Рассматривается механизм распространия через уязвимости в распространенном ПО (уязвимости типа «memory corruption»)

Эксплуатация уязвимости

- Шеллкод – набор машинных инструкций, осуществляющих эксплуатирование удаленной уязвимости «memory corruption»



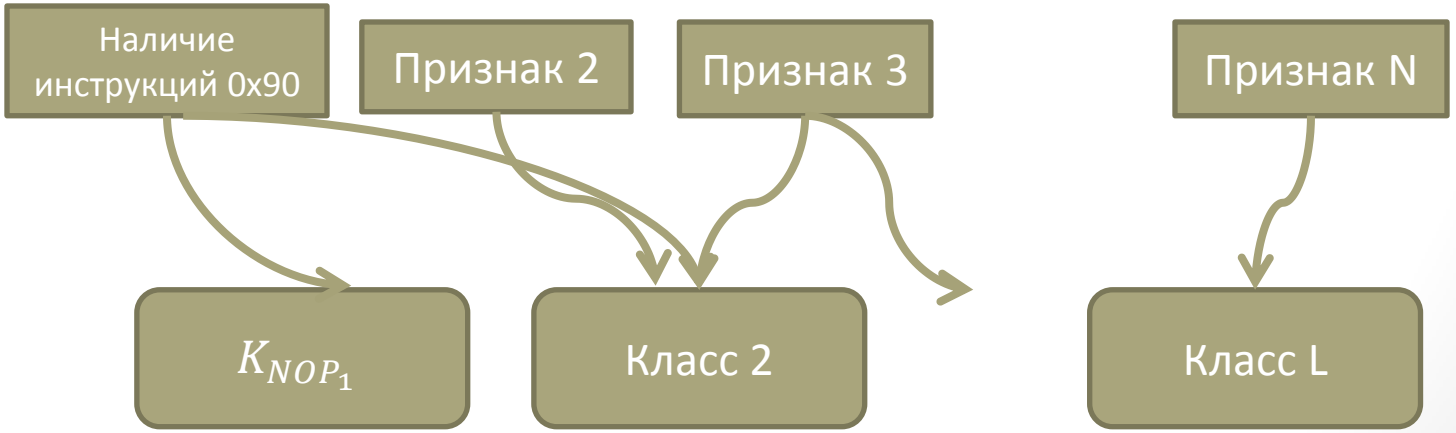
Классификация шеллкодов

Пусть заданы наборы $Leg = \{l_1, \dots, l_k\}, Mal = \{m_1, \dots, m_n\}$

Пространство шеллкодов M разбивается на классы $K_1, \dots, K_l$ по набору признаков вредоносных инструкций:

$$M = \bigcup_{i=1}^{l} K_i$$

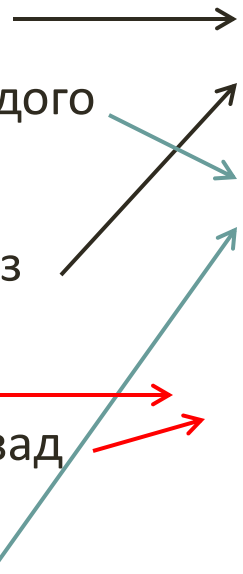
При этом $Mal = \bigcup_{i=1}^{l} Mal(K_i)$ и $Leg \neq \bigcup_{i=1}^{l} Leg(K_i)$ в общем случае.



Классификация шеллкодов

Признаки

- Корректное дизассемблирование байтового потока с каждого смещения
- Наличие цепочки определенной длины из инструкций 0x90
- Наличие GetPC кода
- Условные переходы назад
- Условные переходы в цепочке определенной длины в один адрес
- …



Классы шеллкодов

- Содержащие простейший NOP-след
- Содержащие батутный NOP-след
- Самораспаковывающийся шеллкод
- Метаморфный шеллкод
- … (всего выделено 19 классов)

Классификация методов обнаружения шеллкодов

- Статические
- Динамические
- Гибридные

Анализ методов

Статические

- Возможно полное покрытие кода программы
- В общем случае быстрее динамических

- Задача обнаружения метаморфного шеллкода неразрешима
- Задача обнаружения полиморфного шеллкода NP-полна

Динамические

- Устойчивы к обфускации

- Требуют больше накладных расходов
- Покрытие программы не полно
- Сложность эмуляции окружения
- Существуют техники обнаружения выполнения программы в виртуальном окружении

Результаты обзора методов

- Ни один из существующих методов обнаружения шеллкодов не обеспечивает полного покрытия классов $K_1 \dots K_l$ шеллкодов;
- Методы, имеющие низкую вычислительную сложность, характеризуются высокой долей ложных срабатываний;
- Методы, имеющие высокую точность, характеризуются высокой вычислительной сложностью.

Актуальна разработка комбинированного метода, который позволит снизить вероятность ложных срабатываний при одновременном уменьшении вычислительной сложности по сравнению с простой комбинацией существующих методов

Задача работы

Требуется разработать комбинированный алгоритм обнаружения шеллкодов в высокоскоростных каналах такой, что:

- вероятность ложных срабатываний алгоритма минимальна;
- вычислительная сложность алгоритма минимальна по сравнению с простой комбинацией алгоритмов;
- обеспечивается полное покрытие классов шеллкодов;
- учитывается профиль трафика в канале;
- учитывается директивный интервал анализа трафика.

(Задача многокритериальной оптимизации)

Разработка гибридного классификатора

Задача разделяется на подзадачи:

1. Классификация пространства шеллкодов (выделено 19 классов)
2. Построение библиотеки элементарных классификаторов – построение набора алгоритмов, обнаруживающих специфичные классы шеллкодов
3. Алгоритм построения классификатора – решение оптимизационной задачи генерации графа из элементарных классификаторов
4. Алгоритм выполнения классификатора (теория построения расписаний)

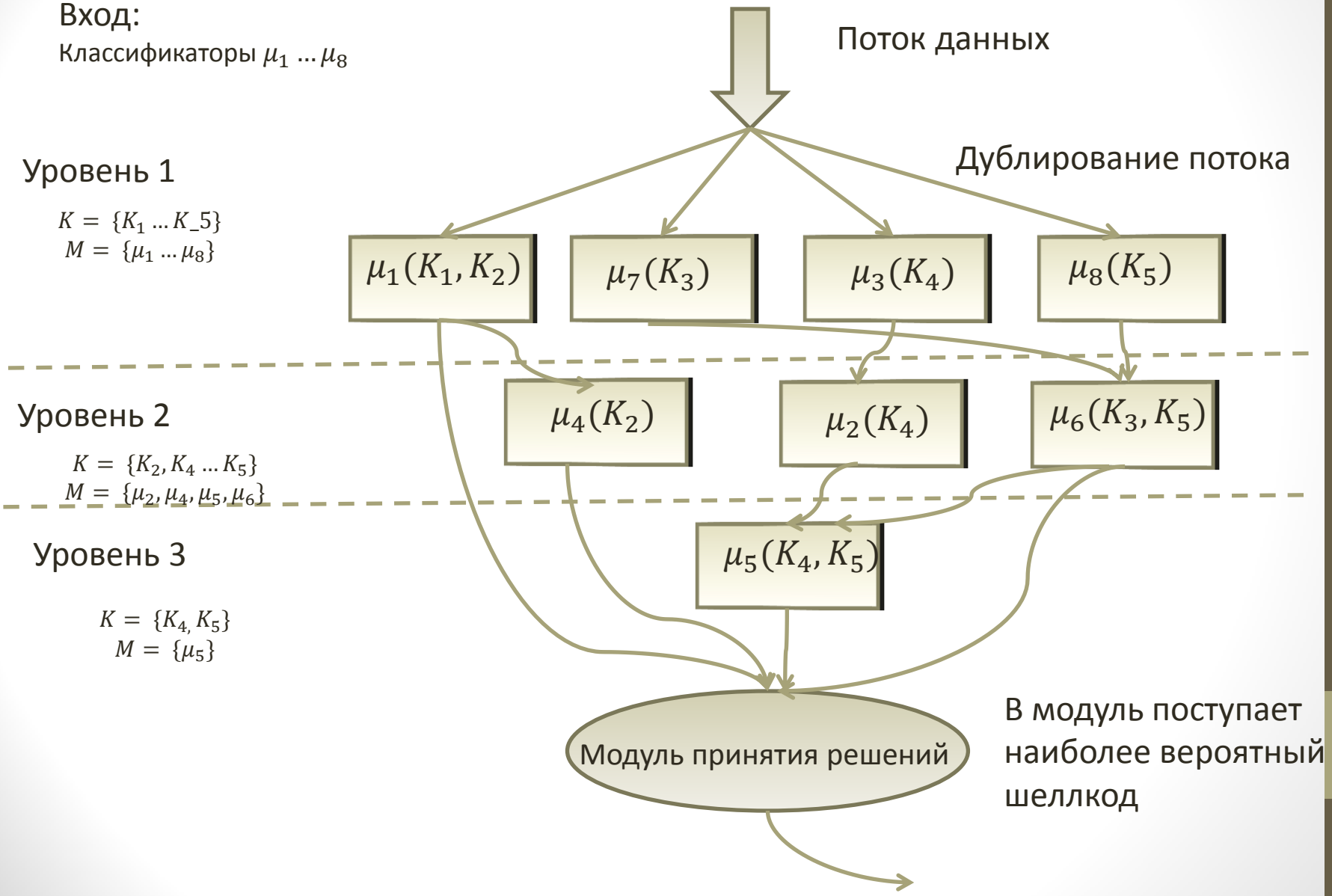
Переборный алгоритм построения топологии графа

- Вход: $M = \{\mu_i(fn_i, fp_i, c_i)\}$ - множество классификаторов с вероятностью ошибки первого и второго рода *fn* и *fp* и сложностью *C*
- Выход: искомый граф

1. Инициализация начальной вершины
2. Выделение множества классов шеллкодов *K*, определяемых классификаторами *M*
3. Построение уровня классификатора, обеспечивающего полное покрытие обнаруживаемых классов *K* и оптимального в терминах *fn*, *fp*, *C*
4. Связывание классификаторов построенного уровня с классификаторами предыдущего уровня по обнаруживаемым классам шеллкодов
5. Исключение из множества *M* классификаторов построенного уровня
6. Если множество *M* не пусто, вернуться на шаг 2

Гибридный классификатор

Вход:
Классификаторы $\mu_1 \dots \mu_8$



В модуль поступает наиболее вероятный шеллкод

Результаты тестирования

Гибридный классификатор по сравнению с линейной комбинацией классификаторов

- На вредоносных данных время работы и вероятность ложных срабатываний совпадают

	Гибридный кл.	Л/К
Количество проанализированных трасс	5544	5544
Количество обнаруженных шеллкодов	5352	5352
Время работы (сек)	55.41	54.88

- На легитимных данных скорость работы гибридного классификатора в **2,5** раза выше

	Гибридный кл.	Л/К
Количество проанализированных трасс	2400	2400
Количество обнаруженных шеллкодов	121	88
Время работы (сек)	11.6	26.75

Спасибо за внимание!