Способы вычисления меры сходства событий безопасности для процесса корреляции

Федорченко Андрей

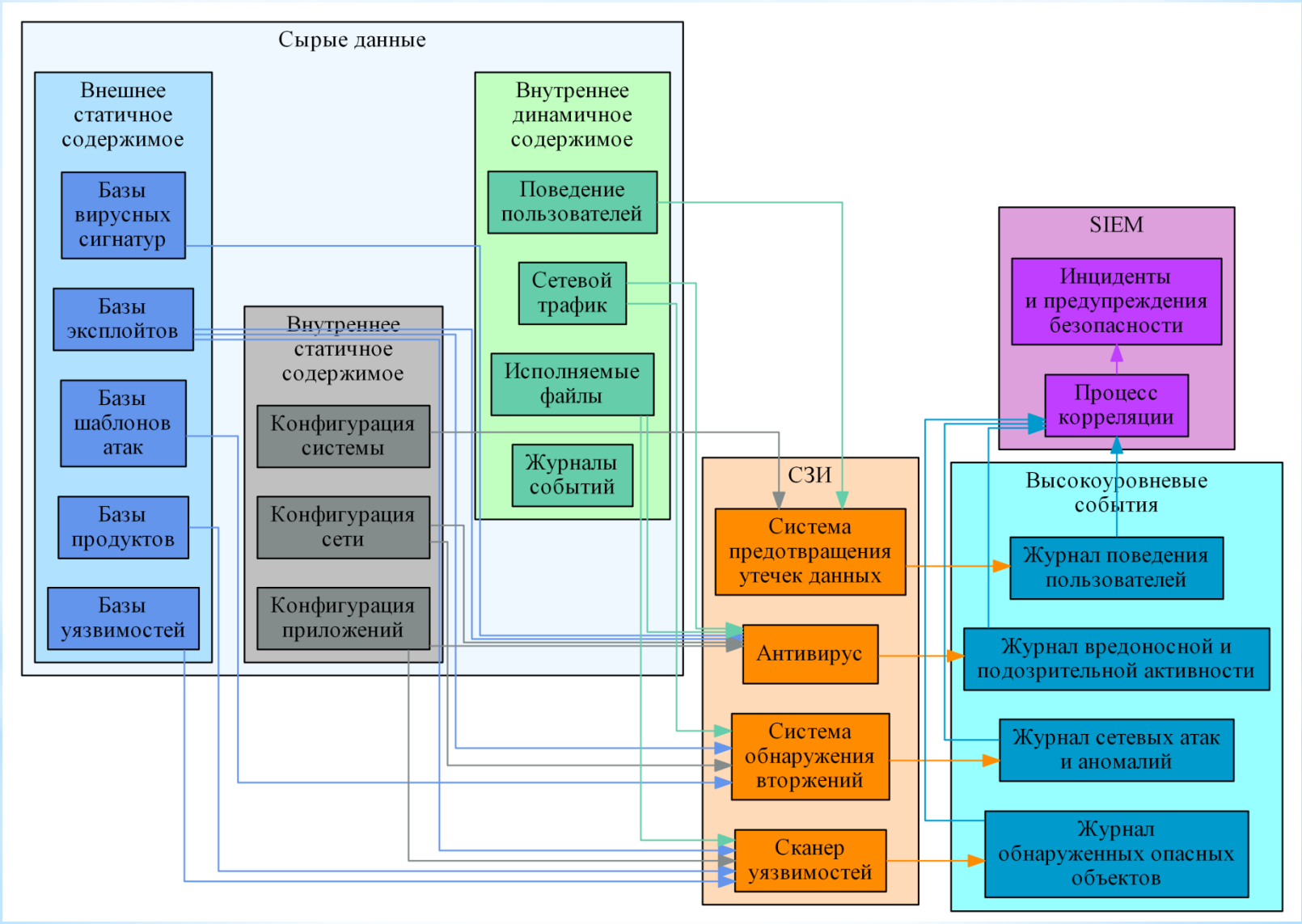
Университет ИТМО
СПИИРАН

РусКрипто'2018
Солнечногорск, 20-23 марта 2018

Место и роль процесса корреляции в SIEM-системах

- Выявление связей между разнородными и разноуровневыми событиями;
- Обнаружение вредоносной, атакующей и аномальной активности;
- Определение источника и цели атаки;
- Обнаружение многошаговых атак.

Входные данные для процесса корреляции (1/3)



Входные данные для процесса корреляции (2/3)

Формальное описание событий безопасности

$$\{e_1, e_2, \ldots, e_k\} \in E^L, \quad \{t_1, t_2, \ldots, t_n\} \in T^L$$

где *E* – множество событий журнала *L*, а *T* – множество типов событий журнала *L:*

$$\{p_1, p_2, \ldots, p_m\} \quad \in \quad P^T \qquad\qquad E \to T, P$$

где *P* множество свойств множества типов *T.*

Группы свойств событий:

- Свойства, идентифицирующие событие;
- Свойства, идентифицирующие источники события;
- Свойства временной привязки события;
- Свойства, идентифицирующие результат действия (попытки);
- Информационные свойства.

Входные данные для процесса корреляции (3/3)

в ОС Windows 10

EventRecordID	EventID	SystemTime	SubjectUserSid	SubjectUserName	SubjectLogonId	ProcessId	CallerProcessId	TargetUserName	TargetDomainName	…
27263	4624	2018-03-20 22:33:12.476223600Z	S-1-5-18	LAB$	0x3e7	0x28c	NaN	NaN	NaN	…
27262	4672	2018-03-20 22:33:12.011372600Z	S-1-5-18	СИСТЕМА	0x3e7	NaN	NaN	NaN	Nan	…
27258	4798	2018-03-20 22:33:11.222295200Z	S-1-5-18	LAB$	0x3e7	NaN	0x1574	User	LAB	…
27251	4648	2018-03-20 22:33:10.972803200Z	S-1-5-18	LAB$	0x3e7	0x1b08	NaN	UMFD-2	Font Driver Host	…

Входные данные для процесса корреляции (3/3)

в условно-неопределенной инфраструктуре

Свойство_1	Свойство_2	Свойство_3	Свойство_4	Свойство_5	Свойство_6	Свойство_7	Свойство_8	Свойство_9	Свойство_10	… Свойство_n
27263	4624	2018-03-20 22:33:12.476223600Z	S-1-5-18	LAB$	0x3e7	0x28c	NaN	NaN	NaN	…
27262	4672	2018-03-20 22:33:12.011372600Z	S-1-5-18	СИСТЕМА	0x3e7	NaN	NaN	NaN	Nan	…
27258	4798	2018-03-20 22:33:11.222295200Z	S-1-5-18	LAB$	0x3e7	NaN	0x1574	User	LAB	…
27251	4648	2018-03-20 22:33:10.972803200Z	S-1-5-18	LAB$	0x3e7	0x1b08	NaN	UMFD-2	Font Driver Host	…

Задачи процесса корреляции

- Группировка низкоуровневых событий в более высокоуровневые события;
- Определение взаимосвязей между разноуровневыми событиями и информацией безопасности;
- Определение важности событий и их групп в рамках задачи обеспечения безопасности;
- Обнаружение инцидентов и предупреждений безопасности.

Проблемы выполнения задач корреляции

- Высокая степень разнородности данных о безопасности;
- Большое количество источников информации и событий безопасности;
- Ручное и объектно-зависимое конфигурирование процесса корреляции;
- Расширение отраслей применения SIEM-систем;
- Увеличение сложности проводимых атак.

Цель и задачи исследования

Глобальная цель: разработка **адаптивной** методики корреляции событий и информации безопасности в условно-неопределенных инфраструктурах.

Глобальные задачи:

- анализ структур входных данных;
- выявление разноуровневых элементов инфраструктуры и анализ их функциональных связей;
- анализ поведения высокоуровневых объектов;
- анализ и прогнозирование состояния инфраструктуры.

Частная задача: исследование метрического пространства признаков событий безопасности для:

- оцифровки признаков;
- нормировки признаков;
- извлечения полезных признаков.

Методы корреляции

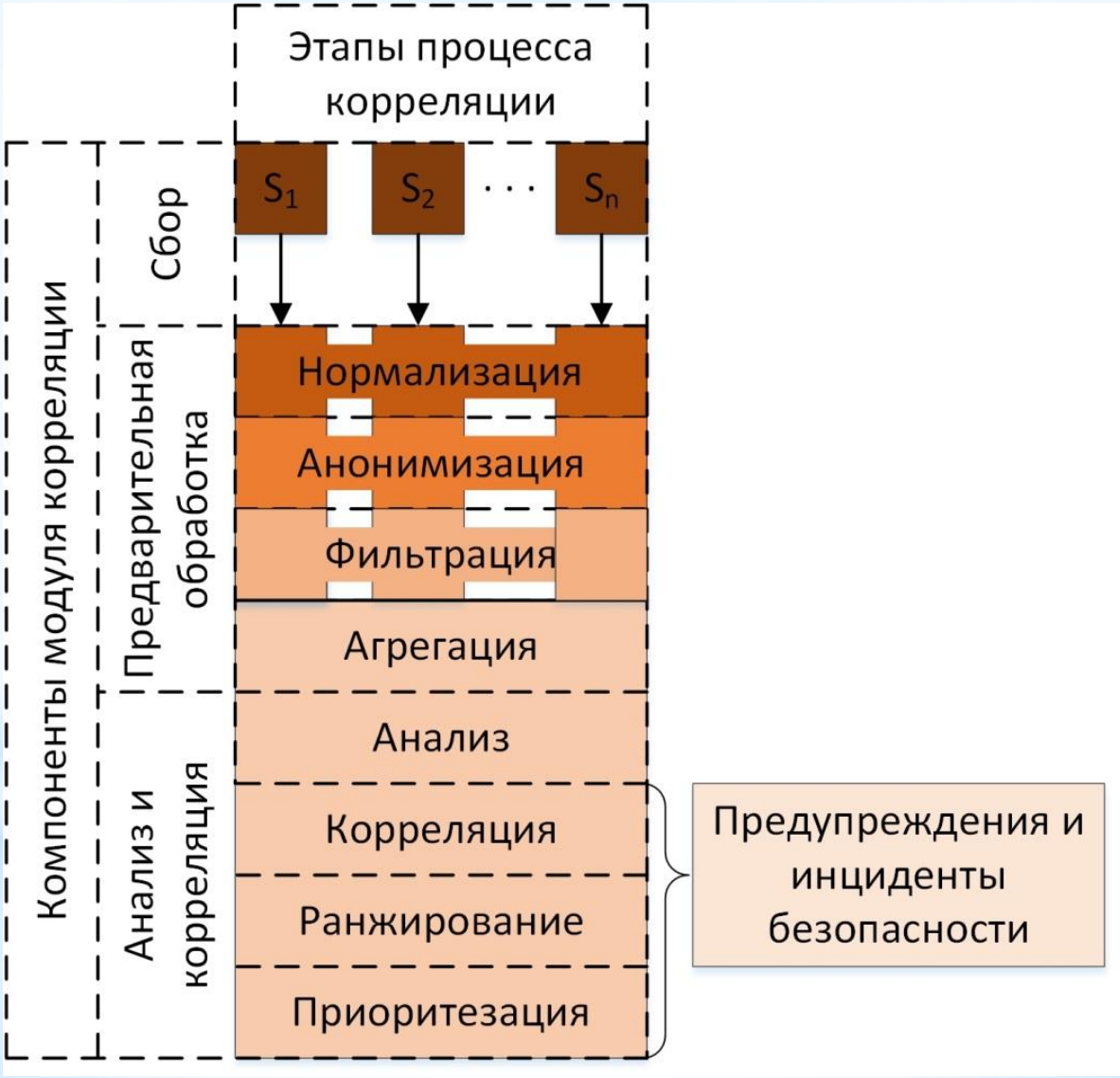
Сигнатурные методы:

- правило-ориентированные **[R. Sadoddin, A. Ghorbani, 2006]**, **[A. Hanemann, P. Marcu, 2008]**, **[T. Limmer and F. Dressler, 2008**];
- шаблонно-ориентированные (на основе сценариев) **[R. Sadoddin, A. Ghorbani, 2006**];
- граф-ориентированные **[A. Muller, 2009]**, **[P. Ning and D. Xu, 2008]**;
- на основе машины конечных автоматов **[A. Muller, 2009]**, **[A.A. Ghorbani et al., 2010]**;
- на основе схожести состояний **[M. A. Hasan, 1999]**, **[U. Zurutuza, R. Uribeetxeberria, 2004]**;
- и другие.

Обучаемые методы:

- байесовские сети **[R. Sadoddin, A. Ghorbani, 2006]**, **[A. Muller, 2009]**, **[D.W. Guerer et al., 1996]**;
- иммунные сети **[A. Muller, 2009]**, **[D.W. Guerer et al., 1996]**;
- искусственные нейронные сети **[A.Muller, 2009]**, **[D.W.Guerer et al., 1996]**, **[H.T. Elshoush and I.M. Osman, 2001]**;
- и другие.

Этапы процесса корреляции



Статистические подходы для корреляции событий безопасности

Вероятностно-статистический:

- неизменность условий воспроизведения событий;
- однотипность событий;
- многократность воспроизведения событий;
- генеральная совокупность – домысливаемая совокупность объектов на основе статистической выборки (наблюдаемых объектов).

Логико-алгебраический:

- отсутствие априорных сведений о вероятностной природе событий;
- «генеральная совокупность» – совокупность наблюдаемых объектов.

Задача группировки событий в семантически связанные наборы:

- отсутствие учителя;
- множественные пропуски в исходных данных (как следствие разнотипности событий);
- преимущественно номинальные типы признаков.

Подходящие методы машинного обучения: **метрические методы** (кластер-анализ, кластеризация).

Критерии схожести

Мера сходства (коэффициент [индекс] сходства, мера ассоциации, метрика) – количественный показатель схожести («близости») сравниваемых объектов, вычисляемый с помощью функции расстояния.

Функция расстояния	Особенности	Недостатки
Минковского	Общий случай для семейства параметрических функций расстояния от p	Вычисление расстояния по признакам только количественного типа
Манхэттенское	$p=1$	
Евклидово	$p=2$	
Чебышева	$p=\infty$	
Махаланобиса	Учитывает зависимость между значениями признаков	
Левенштейна (Дамерау-Левенштейна)	Вычисление расстояния по признакам, типизированным произвольным словарем	Чувствительны к масштабу значений
Джаро (Джаро-Винклера)		

Отношения между свойствами событий (1/4)

Однотипное равнозначное свойство p – это свойство, эквивалентное по типу и смыслу содержимого для двух различных типов событий *t1* и *t2*:

$$\forall p \in P^T : \quad p \in t_1, p \in t_2; \quad t_1, t_2 \in T$$

Однотипные неравнозначные свойства p1 и *p2* – это свойства, эквивалентные по типу содержимого:

$$p_1, p_2 \in P^T : \quad p_1 \sim p_2$$

Разнотипные свойства – это свойства, эквивалентные по значениям содержимого при явной разнице между типами содержимого.

Один тип событий *t* может содержать несколько однотипных неравнозначных и разнотипных свойств *p* в своей структуре:

$$p_1 \sim p_2 \sim \cdots \sim p_s \quad \{p_1, p_2, \ldots, p_s\} \in P^t$$

где *s* – количество однотипных или разнотипных свойств типа *t.*

Отношения между свойствами событий (2/4)

<table>
<tr><th colspan="2">Свойства</th><th colspan="2">Связь</th></tr>
<tr><td rowspan="2">Однотипные</td><td>Равнозначные</td><td colspan="2">Прямая</td></tr>
<tr><td>Неравнозначные</td><td>Однотипная</td><td rowspan="2">Косвенная</td></tr>
<tr><td colspan="2">Разнотипные</td><td>Разнотипная</td></tr>
</table>

Пример связей типов событий в ОС Windows 8

ID типа, значение	Свойство	ID типа, значение	Свойство	Тип связи
4689, «Завершение процесса»	Process Information/ ProcessID	4673, «Вызвана привилегирован-ная служба»	Process Information/ ProcessID	Прямая
4688, «Создание процесса»	Process Information/ NewProcessID	4688, «Создание процесса»	Process Information/ CreatorProcessID	Косвенная однотип-ная
4688, «Создание процесса»	Process Information/ NewProcessID	4688, «Создание процесса»	Process Information/ NewProcessName	Косвенная разнотип-ная

Отношения между свойствами событий (3/4)

Этапы определения косвенной однотипной связи:

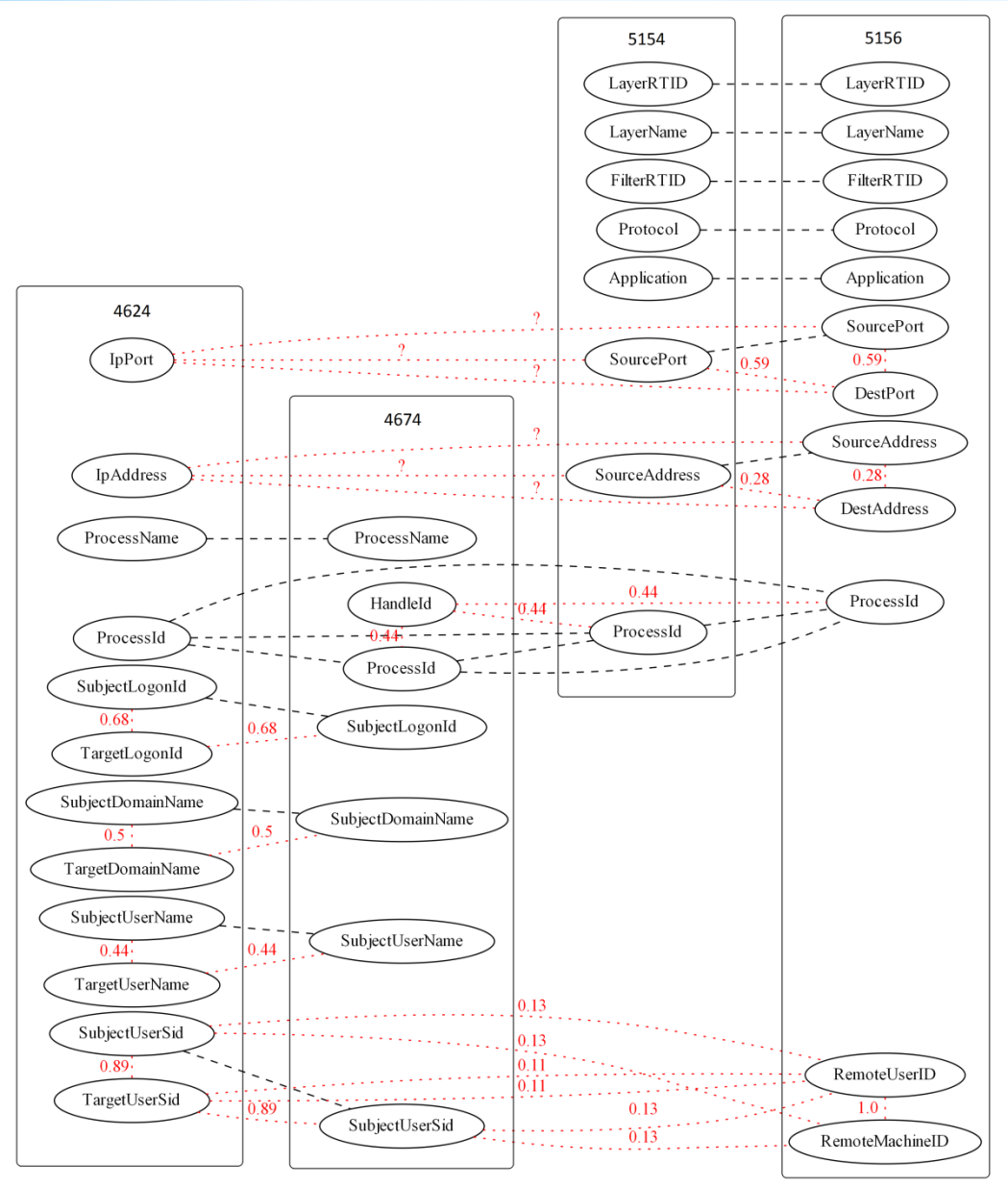
- формирование множеств значений равнозначных свойств событий;
- вычисление показателя схожести неравнозначных свойств - операция выполняется на основе определения пересечения множеств значений свойств и последующей нормировки

$$Sim^{p_1,p_2} = \frac{|V^{p_1} \cap V^{p_2}|}{|V^{p_1}| + |V^{p_2}| - |V^{p_1} \cap V^{p_2}|}, \qquad \{p_1, p_2\} \in P, \qquad |V^{p_1}|, |V^{p_2}| > 0$$

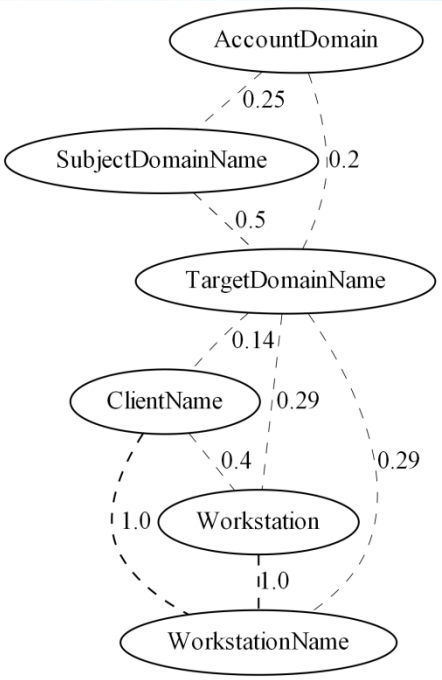
где V^{p_1} и V^{p_2} - множества значений свойств p_1 и p_2 соответственно;

- объединение пар неравнозначных однотипных свойств.

Отношения между свойствами событий (4/4)



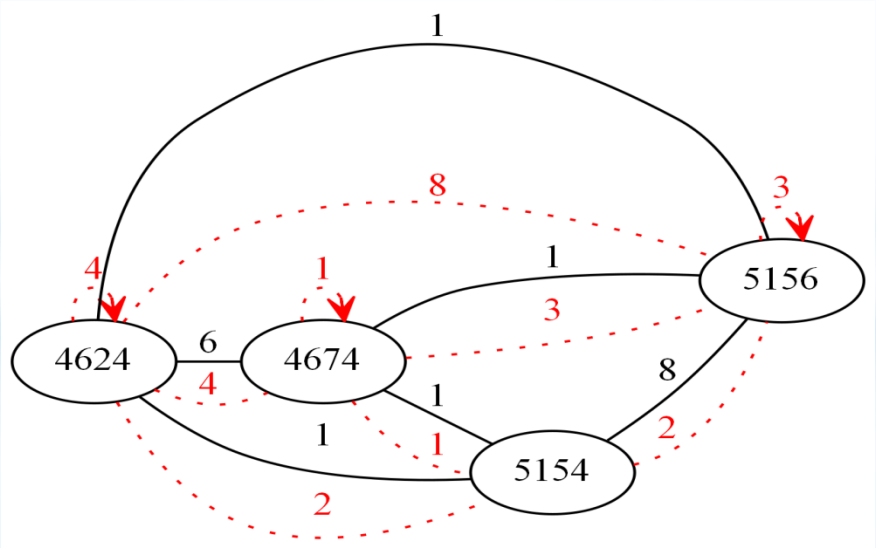
Пример вычисленных показателей схожести однотипных неравнозначных (красным) и однотипных равнозначных (черным) свойств типов событий ОС Windows 8



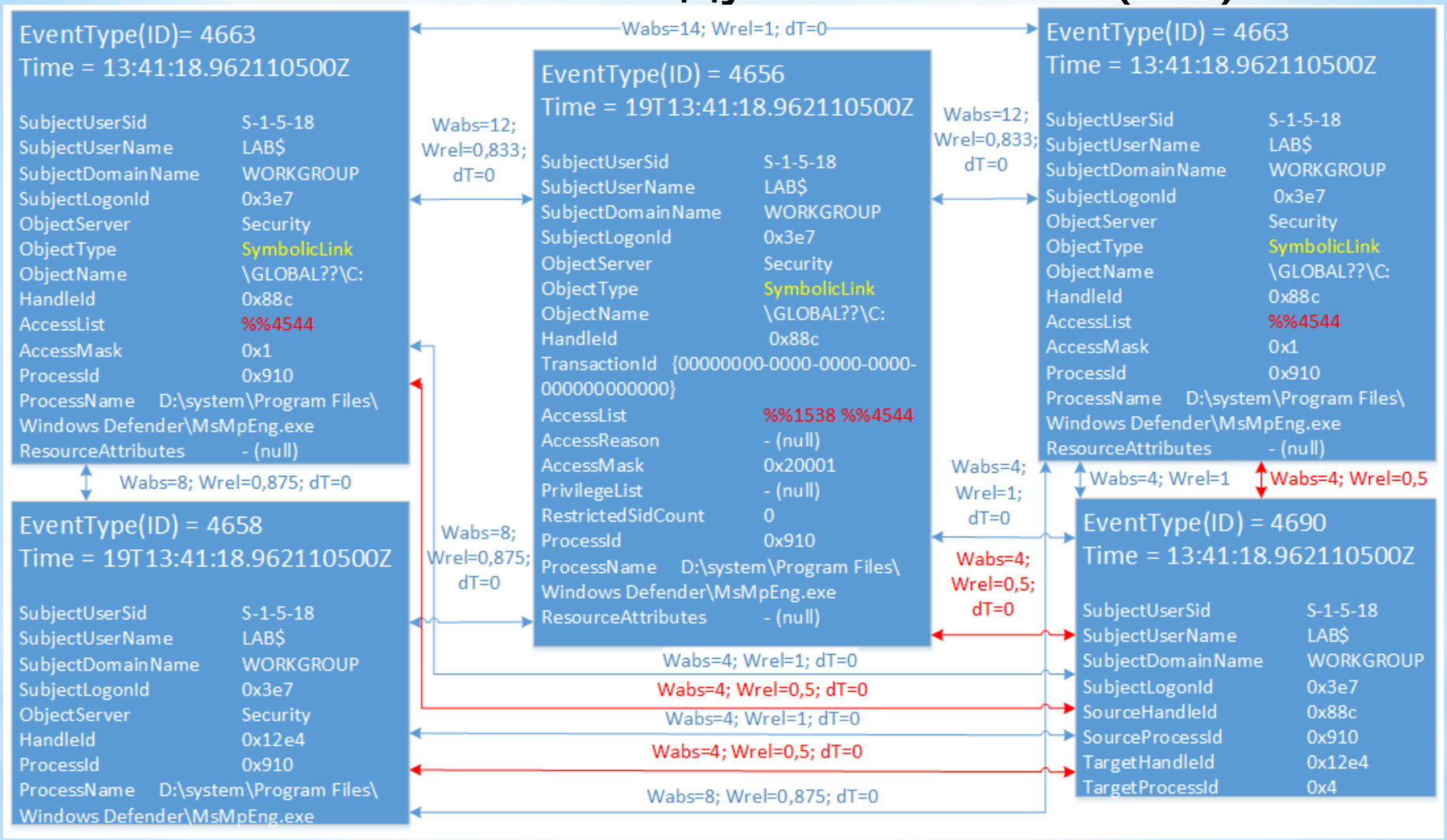
Отношения между событиями (1/2)

Абсолютный вес $w_{abs}^{t_1,t_2}$ между двумя типами событий t_1 и t_2 выражается количеством их сопоставимых (прямо или косвенно) свойств $p_{eq}^{t_1,t_2}$ (где N - функция определения количества элементов): $w_{abs}^{t_1,t_2} = N\left(p_{eq}^{t_1,t_2}\right)$

Относительный вес связи между двумя экземплярами событий определяется количеством сопоставимых свойств, значения которых идентичны друг другу. Данная метрика представляется выражением: $w_{rel}^{t_1,t_2} = \frac{N(p_i^{t_1} == p_i^{t_2})}{n},\ i = 1 \dots n,\ n = w_{abs}^{t_1,t_2},\ p_i^{t_1} \sim p_i^{t_2},$ где $p_i^{t_1} == p_i^{t_2}$ обозначает множество сопоставимых свойств.



Отношения между событиями (2/2)



Пример вычисленных абсолютных весов (W_{abs}) и относительных весов (W_{rel}) по прямым (синим) и косвенным однотипным (красным) связям между типами событий ОС Windows 8

Граф связей типов событий на основе анализа их структур

В результате анализа структур типов событий сформирован неориентированный граф связей *G*

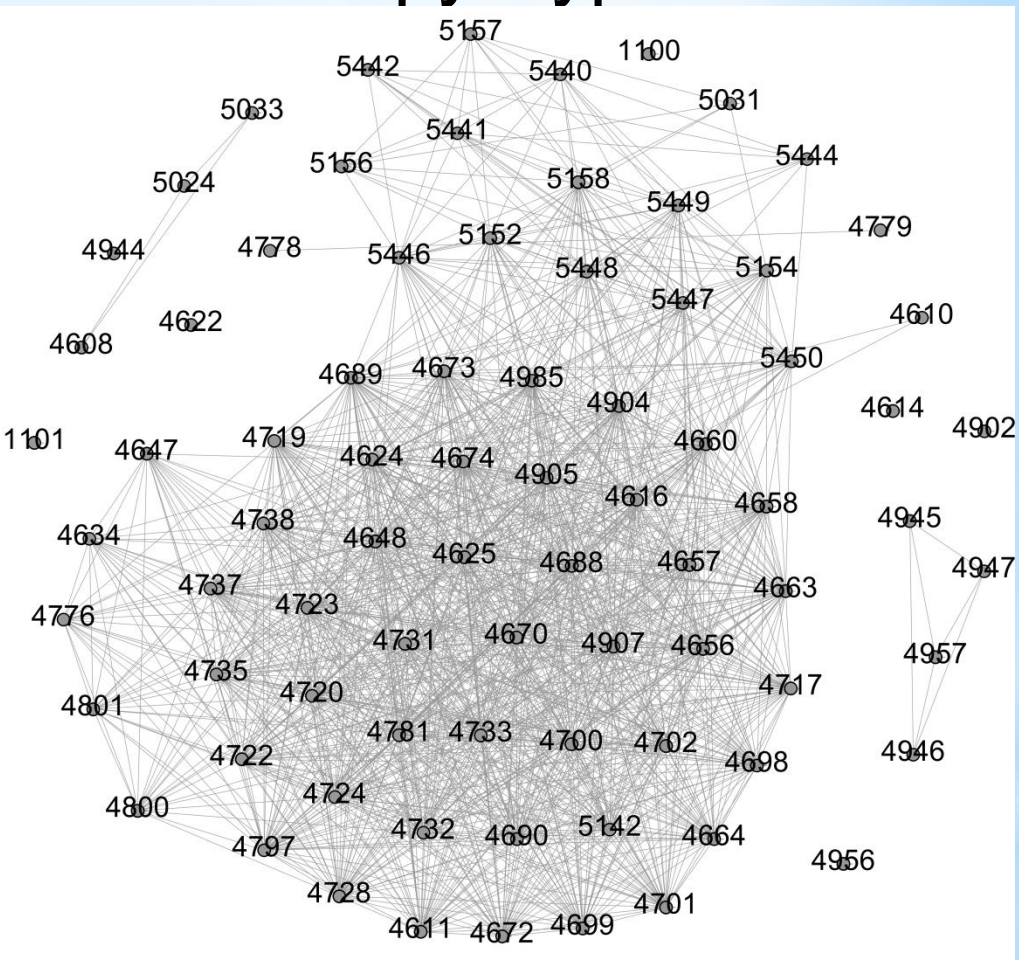
$$G = (T, P, \varphi), \quad \varphi : P \to T \times T$$

На рисунке:

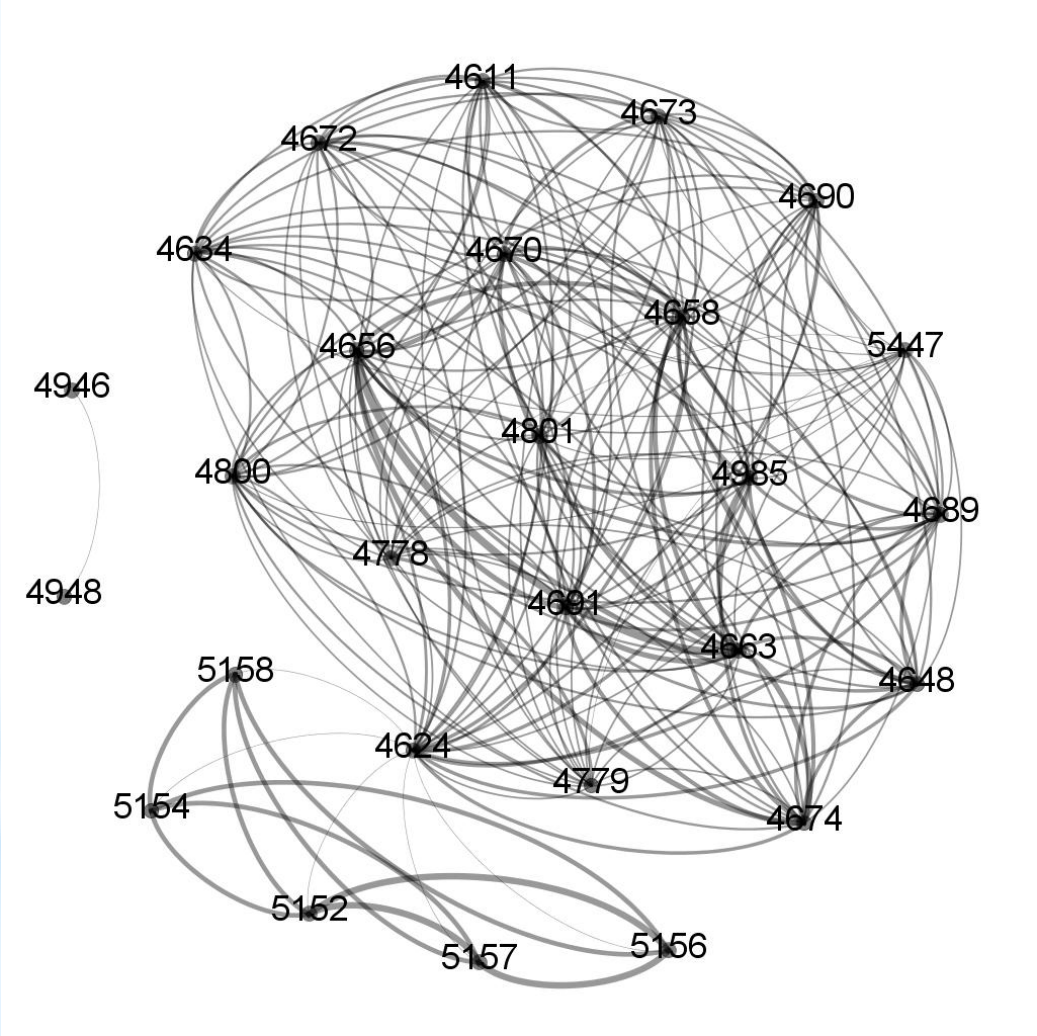
- количество выявленных типов событий - 80 из 418 заявленных в документации;
- свойств - 158,

53 – уникальные (встречаются только в одном типе событий);

89 – смежные (встречаются более чем в одном типе событий).



Уточнение графа связей типов событий



Подведение итогов

Текущие результаты исследования:

- проанализированы структуры типов событий ОС Windows;
- разработана модель связей типов событий;
- исследовано метрическое пространство событий;
- рассмотрены меры сходства на основе прямых и косвенных однотипных связей.

Дальнейшая работа:

- кластер-анализ показателей взаимного использования свойств событий, для сегментации их на уровни;
- расширение (уточнение) пространства признаков;
- кластер-анализ событий безопасности.

СПАСИБО ЗА ВНИМАНИЕ!

Работа выполнена при поддержке Гранта президента Российской Федерации (МК-314.2017.9).

Контактная информация:

Федорченко Андрей
fedorchenko@comsec.spb.ru