МОДЕЛЬ ПРОЦЕССА ОБРАБОТКИ СОБЫТИЙ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ

THE MODEL OF PROCESSING EVENTS OF INFORMATION SECURITY

A. Gaynov
I. Zavodtsev

Summary. This article reviews the existing models of the process of handling information security events, identified and analyzed their strengths and weaknesses. The improved model of this process is proposed, which differs from existing ones that for the events processing are used the special characteristics, contained in log files and data received from employees / customers of the organization, which takes into account the performance requirements. At the same time, the data received from the employees / customers of the organization are represented by the different modalities from the graphic and text materials characterized by different statistical properties. The use of these data in SIEM-systems will improve the speed and accuracy of detection the information security incidents.

Keywords: an information security incident, a SIEM-system, a log-file, multimodality.

Гайнов Артур Евгеньевич
Соискатель, Кубанский институт Информационной защиты, г. Краснодар
ArturGaynov@mail.ru
Заводцев Илья Валентинович
К.т.н., доцент, Кубанский институт информационной защиты, г. Краснодар
nilrs@mail.ru

Аннотация. В настоящей статье рассмотрены существующие модели процесса обработки событий информационной безопасности, выявлены и проанализированы их сильные и слабые стороны. Предложена усовершенствованная модель данного процесса, отличающаяся от существующих тем, что для обработки событий используются признаки содержащиеся, как в лог-файлах, так и в данных, полученных от сотрудников/клиентов организации, с учетом требований производительности. При этом данные, полученные от сотрудников/клиентов организации, представлены разнообразными модальностями из пространства графических и текстовых материалов, характеризующихся различными статистическими свойствами. Использование в SIEM-системах указанных данных будет способствовать повышению оперативности и точности обнаружения инцидентов информационной безопасности.

Ключевые слова: инцидент информационной безопасности, SIEM-система, лог-файл, мультимодальность.

Введение

Информационные технологии (ИТ) сегодня приобрели глобальный трансграничный характер и стали неотъемлемой частью всех сфер деятельности личности, общества и государства. Обеспечение бесперебойного функционирования информационной инфраструктуры (ИТИ), в первую очередь критической, является одним из национальных интересов Российской Федерации (РФ) в информационной сфере [1].

Вместе с тем, расширение областей применения ИТ, являясь фактором развития, одновременно порождает новые угрозы информационной безопасности (ИБ). Одним из негативных факторов, влияющих на состояние ИБ, является наращивание рядом зарубежных стран возможностей информационно-технического воздействия на ИТИ [1].

На сегодняшний день к числу приоритетных направлений совершенствования системы защиты информации, обрабатываемой в ИТИ, относится развитие системы мониторинга инцидентов ИБ (ИИБ) или, иначе говоря, SIEM-систем [3].

Релевантные работы

Одной из основных целей функционирования SIEM-систем является обнаружение ИИБ за счет эффективной обработки событий ИБ (СИБ), в частности, определение как относящихся или не относящихся к ИИБ [2]:

$ISE = \{ise_1, \ldots, ise_i\}$ — множество СИБ. Каждое *СИБ* $ise \in ISE$ представляет собой последовательность признаков

$$W_{ise} = (w_1, \ldots, w_{iise}),$$

где i_{ise} — длина СИБ ise;

$ISI = \{isi_1, \ldots, isi_j\}$ — конечное множество меток классов ИИБ;

$y^*: ISE \rightarrow ISI$ — неизвестная целевая зависимость.

Для оценки эффективности функционирования SIEM-систем используются критерии: производительность, оперативность, точность.

Под производительностью понимается отношение количества обработанных СИБ в установленный интервал времени:

$$\mathrm{ESP} = \frac{\mathrm{SUM}_{\mathrm{ISE}}}{\text{”}\,\mathrm{t}_{\mathrm{ISE}}}, \quad (1)$$

где SUM_{ISE} — количество обработанных СИБ, $\Delta \mathrm{t}_{\mathrm{ISE}}$ — интервал времени, в течении которого производилась обработка СИБ.

Под оперативностью понимается интервал времени, необходимый для обнаружения одного ИИБ:

$$\Delta \mathrm{t} = \mathrm{t}_2 - \mathrm{t}_1, \mathrm{t}_2 > \mathrm{t}_1, \quad (2)$$

где t_1 — время обнаружения СИБ, t_2 — время подтверждения того, что СИБ является ИИБ.

Под точностью понимается доля СИБ, принадлежащих данному ИИБ, относительно всех СИБ, которые SIEM-система отнесла к данному ИИБ.

$$\mathrm{Presicion} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}, \quad (3)$$

где, TP — истинно-положительное решение, FP — ложно-положительное решение.

В современных SIEM-системах используются СИБ, полученные из лог-файлов, а также для повышения эффективности обработки СИБ дополнительно применяются данные об ИТИ и состоянии ее защищенности [5]:

$ISE_{PI} = \{ise_{PI1}, \ldots, ise_{PIi}\}$ — множество СИБ, полученных из лог-файлов. Каждое СИБ $ise_{PI} \in ISE_{PI}$ представляет собой последовательность признаков

$$\mathrm{W}_{\mathrm{ise}_{\mathrm{PI}}} = \left(\mathrm{w}_{1\mathrm{PI}}, \ldots, \mathrm{w}_{\mathrm{iise}_{\mathrm{PI}}}\right),$$

где $\mathrm{i}_{\mathrm{ise}}$ — длина СИБ ise_{PI}.

$INV = \{inv_1, \ldots, inv_n\}$ — множество защищаемых ресурсов (ЗР) ИТИ. Каждый ЗР $inv \in INV$ представляет собой последовательность данных

$$\mathrm{S}_{\mathrm{inv}} = (\mathrm{w}_1, \ldots, \mathrm{w}_{\mathrm{n}_{\mathrm{inv}}}),$$

где n_{inv} — длина $ЗР_{inv}$;

$VUL = \{vul_1, \ldots, vul_m\}$ — множество уязвимостей (УЗ) ИТИ. Каждая УЗ $vul \in VUL$ представляет собой последовательность данных

$$\mathrm{J}_{\mathrm{vul}} = (\mathrm{j}_1, \ldots, \mathrm{j}_{\mathrm{m}_{\mathrm{vul}}}),$$

где m_{vul} — длина $УЗ$ vul.

В данном подходе в результате функционирования SIEM-системы определяются не только появление одного или нескольких СИБ, с которыми связана значительная вероятность создания угрозы ИБ, но и ЗР, на которые может быть направлена обнаруженная угроза ИБ, и УЗ, которые могут быть использованы для ее реализации.

Вместе с тем, современные SIEM-системы имеют и ограничения. Так SIEM-система может получать в качестве входных данных СИБ от любых источников. Следовательно, ничто не помешает злоумышленнику сгенерировать поток ложных СИБ, притворившись одним из сенсоров:

$$\mathrm{ISE}' = \left\{\mathrm{ise}'_1, \ldots, \mathrm{ise}'_{\mathrm{k}}\right\},$$

— множество ложных СИБ. Каждое СИБ ise′ ∈ ISE′ представляет собой последовательность симптомов

$$\mathrm{W}_{\mathrm{ise}'} = \left(\mathrm{w}_1, \ldots, \mathrm{w}_{\mathrm{iise}'}\right),$$

где $\mathrm{i}_{\mathrm{ise}'}$ — длина СИБ ise′.

Наличие подобного потока СИБ значительно ухудшает качество процесса обработки СИБ и может привести к обнаружению ИИБ, которых не существует.

Подобные современные SIEM-системы в общем виде всегда содержат ряд основных подсистем: сбора данных, обработки, хранения, представления [4].

На протяжении последних 20 лет активно исследуется вопрос представления подсистемы обработки в виде процесса, включающего последовательно или параллельно выполняемые этапы [5–11].

За указанное время предложены различные схемы, описывающие процесс обработки СИБ, который можно представить следующим образом:

$$PROC_{ISE} = <ISEPI, INV, VUL, ISI, PR_{ISE}, I_{ISE}>, \quad (4)$$

где PR_{ISE} — база данных правил обработки СИБ, I_{ISE} — интерпретатор, представленный набором этапов обработки СИБ (таблица 1).

Несомненным достоинством модели, представленной в работе [8], является возможность добавления статистических методов, несмотря на то, что она основана на строго детерминированных подходах. Недостатки представленной модели заключаются в отсутствии среди указанных задач элемента предупреждения ошибок и элемента предупреждения нарушений.

Таблица 1.

Авторы моделей	Основные этапы обработки СИБ
Jakobson U. и др. [8]	I_{ISE} = <Com, Ac, Sup, LR, Syn>, где Com — сжатие, Ac — счет, Sup — подавление, LR — логическая замена, Syn — обобщение
Zurutuza U. и др. [12]	I_{ISE} = <Pp, AW, CA, MSC, Cl, Del, Red, Mer>, где Pp — предобработка, AW — анализ предупреждений, CA — корреляция предупреждений, MSC — измерение схожих признаков, Cl — кластеризация, Del — удаление, Red — редукция, Mer — слияние
Sadoddin R. и др. [10]	I_{ISE} = <Nor, Ag, C, SFP, AAS, Pr>, где Nor — нормализация, Ag — агрегация, C — корреляция, SFP — отсеивание ложных срабатываний, AAS — анализ стратегии атаки, Pr — приоритезация
Dadkhah S. и др. [6]	I_{ISE} = <Res, PSA, PC, MSI, F>, где Res — подобие, PSA — предопределение сценариев атак, PC — многоуровневые вычисления на базе предпосылок и последствий, MSI — использование множества источников информации, F –фильтрация
Kruegel C. и др. [9]	I_{ISE} = <Nor, Pp, AsW, VerW, RSpA, RSeA, DSTA, MSC, AIm, Pr>, где Nor — нормализация, Pp — предобработка, AsW — объединение предупреждений, VerW — верификация предупреждений, RSpA — восстановление хода атаки, RSeA — восстановление сессии атаки, DSTA — определение источника и цели атаки, MSC — многошаговая корреляция, AIm — анализ воздействия, Pr — приоритезация
Elshoushand H.T. [7]	I_{ISE} = <Nor, Pp, Pr, VerW, MerW, DSTA, DelEv, MSC, AIn, AIm>, где Nor — нормализация, Pp — предобработка, Pr — приоритезация, VerW — верификация предупреждений, MerW — слияние предупреждений, DSTA — определение источника и цели атаки, DelEv — удаление не коррелируемых событий, MSC — многошаговая корреляция, AIn — анализ намерений, AIm — анализ воздействий
Котенко и др. [5]	I_{ISE} = <Nor, Pp, Anon, AgF, RSpA, RSeA, DSTA, MSC, ADam, Pr, FilRAN>, где Nor — нормализация, Pp — предобработка, Anon — анонимизация, AgF — агрегация и фильтрация, RSpA — восстановление хода атаки, RSeA — восстановление сессии атаки, DSTA — определение источника и цели атаки, MSC — многошаговая корреляция, ADam — анализ ущерба, Pr — приоритезация, FilRAN — фильтрация на основе ранжирования

В [12] производится обзор работ в области обработки СИБ, генерируемых системами обнаружения вторжений. В частности, рассматриваются этапы и операции процесса обработки СИБ, описывается модель данных формата обмена сообщениями, а также приводится пример процесса обработки СИБ для обнаружения типовой атаки. Стоит отметить, что процесс обработки СИБ представлен 3 этапами. В результате выполнения каждого из этапов формируются простые события, мета-события и сценарии атак, по окончании — отчет.

Модель, представленная в [10], в отличие от аналогичных работ точно связывает этапы процесса обработки СИБ с используемыми в них конкретными методами.

В [6] авторы выделяют пять подходов к процессу обработки СИБ. Первый подход заключается в вычислении величины подобия двух СИБ на основе атрибутов, ассоциируемых с этими событиями. СИБ, величина подобия которых достаточно велика, группируются. Второй подход заключается в объединении в последовательность связанных этапов проведения атак на основе заранее определенных шаблонов сценариев атак. Данный подход применяется для получения агрегированного и более высокоуровневого взгляда на угрозы ИБ. Третий подход основывается на формировании сценариев атак путем связывания отдельных этапов их проведения при условии, что один из этапов является необходимым условием для проведения другого. Четвертый подход направлен на приоритезацию и классификацию потоков СИБ в зависимости от источника данных. Пятый подход основан на удалении из процесса обработки событий по заранее определенным правилам (фильтрам). Решение об удалении СИБ из процесса корреляции принимается на основе значений одного или нескольких его атрибутов.

В [9] раскрывают методы и подходы к обработке СИБ в зависимости от фазы процесса, указываются их достоинства и недостатки, а также спорные моменты. В работе процесс обработки СИБ представлен в виде этапов, которые преобразуют оповещения сенсоров в отчеты о вторжениях и направлены на разные аспекты процесса обработки СИБ. Рассматриваются методы обработки СИБ в системах обнаружения вторжений, которые также применимы в SIEM-системах.

В [7] процесс обработки СИБ представлен в виде логических блоков. Отличительной особенностью данной работы является представление модели процесса обработки СИБ, которая снижает количество обрабатываемых событий так рано, как это только возможно. Это осуществляется путем вывода из процесса обработки СИБ незначимых или ложных событий еще на на-

чальных этапах процесса. Важно отметить, что данная модель не лишена недостатков. Во-первых, в блоке фильтрации будут обрабатываться, в том числе, и дубликаты СИБ, так как этап слияния предупреждений находится на более высоком уровне. Во-вторых, этап удаления из процесса обработки СИБ данных, которые не могут быть обработаны, не оставляет процессу права на ошибку.

В [5] процесс обработки СИБ является сложной задачей и разбит на подзадачи с помощью декомпозиции. Применение декомпозиции к системе, реализующей процесс обработки СИБ, позволяет представить систему в виде простых функциональных модулей. Такой подход обеспечивает рассмотрение каждого модуля независимо друг от друга. При этом четко определены функциональная нагрузка каждого модуля и порядок их взаимодействия. Кроме того, декомпозиция системы, реализующей процесс обработки СИБ, упрощает разработку, отладку и тестирование отдельных модулей. Важно отметить, что в рамках модульного представления процесса обработки СИБ не отражена реализация процесса верификации или проверки источников СИБ на подлинность.

Несмотря на наличие различий SIEM-систем по процессу обработки СИБ, их объединяет одно общее ограничение: в качестве источников СИБ используются исключительно лог-файлы аппаратных и программных элементов, образующих ИТИ (серверы, компьютеры, коммутаторы, системы обнаружения атак, антивирусные средства и т.д.) [3].

Вместе с тем, для обнаружения СИБ могут и должны обрабатываться признаки, содержащиеся, как в лог-файлах, так и в данных, полученных от сотрудников/клиентов организации [2].

Таким образом, в ходе проведенного анализа выявлено противоречие: с одной стороны, должны быть обнаружены и обработаны все СИБ в режиме реального времени или близком к нему, с другой стороны, для обнаружения СИБ в существующих SIEM-системах анализ данных, полученных от сотрудников/клиентов организации, для обнаружения в них признаков СИБ и их последующей обработки не проводится.

Учитывая сказанное, а также в условиях, когда время совершения ИИБ занимает всего несколько секунд или минут, а время обнаружения и обработки СИБ может занимать недели и месяцы, одним из путей разрешения указанного противоречия является разработка научно-методического аппарата обработки СИБ, признаки о которых содержатся, как в лог-файлах, так и в данных, полученных от сотрудников/клиентов организации, с учетом требований оперативности, производительности и точности.

Модель процесса обработки СИБ

Обработка СИБ, признаки о которых содержатся в данных, полученных от сотрудников/клиентов организации, создает целый ряд барьеров.

Это связано с тем, что признаки таких СИБ представлены разнообразными модальностями, полученными из пространства графических и текстовых материалов (текстовое сообщение, скриншот экрана, фотография нарушения и т.п.), характеризующихся различными статистическими свойствами [10]: ISEUs = (P1, P2), где P1 = {p1,…, pk} — множество СИБ, признаки которых выявлены из текстовых обращений, P2 = {p1,…, pq} — множество СИБ, признаки которых выявлены из графический обращений.

Кроме того, данные признаки являются слабоструктурированными, зашумленными и могут иметь пропущенные значения.

Следовательно, модель процесса обработки СИБ требует уточнения и может быть представлена следующим образом (рис. 1):

$$PROC_{ISE} = <ISE_{Pl}, ISE_{Us}, INV, VUL, ISI, PR_{ISE}, I_{ISE}> \quad (5)$$

Основываясь на модели, представленной в работе Котенко [4], интерпретатор включает следующие основные этапы обработки СИБ:

$$I_{ISE} = <Nor, Pp, Pp_{Us}, Anon, AgF, Iden, Ver, RSpA, RSeA, \text{DSTA, MSC, ADam, Pr, Fil}_{RAN}>, \quad (6)$$

где Pp_{Us} — предобработка, Iden — распознавание, Ver — верификация.

Предобработка. Данные, полученные от сотрудников/клиентов организации, необходимо перевести в вещественное пространство признаков. Например, текст можно представить в виде дискретных разреженных векторов с определенным количеством слов, в то время как изображения — с использованием пиксельных интенсивностей.

Распознавание. На данном этапе необходимо из данных, полученных от сотрудников/клиентов организации, извлечь признаки СИБ. Далее следует определить исходное множество правил обработки, в которых присутствуют СИБ, описываемые полученными признаками.

Верификация. После завершения предыдущего этапа необходимо проверить правильность определе-

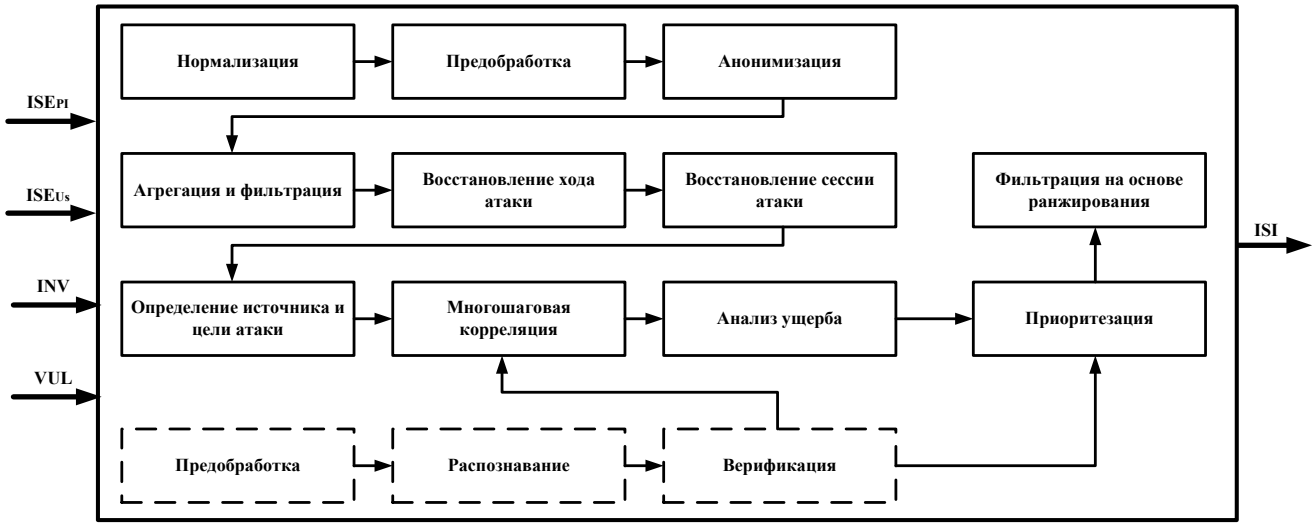

Рис. 1. Функциональная модель процесса обработки СИБ

ния множества правил обработки СИБ. Для указанной проверки предлагается последовательно проводить оценку приоритетности, новизны и специфики правил, использовать фоновые условия, определяющие обстоятельства, влияющие на учет (рассмотрение) тех или иных признаков СИБ, а также определить глубину анализа, устанавливающую временной интервал, в течение которого собираются данные о СИБ. Результатом данного этапа будет окончательное действующее бесконфликтное множество правил, в шаблонах которых имеются СИБ, представленные признаками содержащимися, как в лог-файлах, так и в данных, полученных от сотрудников/клиентов организации.

При этом целевую функцию SIEM-системы можно записать следующим образом:

$$y = f\,(ISE_{PI},\ INV,\ VUL,\ ISE_{Us}), \tag{7}$$

где при $y \to min, ESP \to max, \Delta t \to 0, Presicion \to 1$, $ISE'_{PI} \subset ISE_{PI}$, $ISE'_{Us} \subset ISE_{Us}$.

Заключение

Таким образом, в данной статье предложена усовершенствованная модель указанного процесса, отличающаяся от существующих тем, что для обработки СИБ используются признаки содержащиеся, как в лог-файлах, так и в данных, полученных от сотрудников/клиентов организации, с учетом требований производительности. Использование в SIEM-системах указанных данных будет способствовать повышению оперативности и точности обнаружения ИИБ.

ЛИТЕРАТУРА

1. Российская Федерации. Указ Президента Российской Федерации 2016 г. № 646. Доктрина информационной безопасности Российской Федерации.
2. Национальный стандарт Российской Федерации ГОСТ Р ИСО/МЭК 18044–2007 «Информационная технология. Методы и средства обеспечения безопасности информации. Менеджмент инцидентов информационной безопасности».
3. Котенко И.В., Саенко И. Б., Юсупов Р. М. Новое поколение систем мониторинга и управления инцидентами безопасности / Котенко И. В., Саенко И. Б., Юсупов Р. М. // Научно-технические ведомости СПбГПУ.— 2014.— № 3 (198).— с. 7–18.
4. Котенко И.В., Саенко И. Б., Полубелова О. В., Чечулин А. А. Применение технологии управления информацией и событиями безопасности для защиты информации в критически важных инфраструктурах / Котенко И. В., Саенко И. Б., Полубелова О. В., Чечулин А.А // Труды СПИИРАН.— 2012.— № 1 (20).— с. 27–56.
5. Федорченко А.В., Левшун Д. С., Чечулин А. А., Котенко И. В. Анализ методов корреляции событий безопасности в SIEM-системах. Часть 1 / Федорченко А. В., Левшун Д. С., Чечулин А. А., Котенко И. В. // Труды СПИИРАН.— 2016.— № 4(47).— с. 5–27.
6. Dadkhah S., Shoja M., Taheri H. Alert Correlation through a Multi Components Architecture // International Journal of Electrical and Computer Engineering (IJECE). 2013. vol. 3. no. 4. pp. 461–466.
7. Elshoushand H.T., Osman I. M. An improved framework for intrusion alert correlation // Proceedings of World Congress on Engineering 2012 (WCE2012). 2012. vol. 1. pp. 518–524.
8. Jakobson G., Weissman M. D. Alarm correlation // IEEE Network. 1993. vol. 7(6). pp. 52–59.

9. Kruegel C., Valeur F., Vigna G. Intrusion Detection and Correlation: Challenges and Solutions // University of California, Santa Barbara, USA: Springer. 2005. pp. 29–33.
10. Sadoddin R., Ghorbani A. Alert Correlation Survey: Framework and Techniques // Proceedings of 2006 International Conference on Privacy, Security and Trust: Bridge the Gap Between PST Technologies and Business Services (PST'06). 2006. Article no. 37.
11. Srivastava N., Sahakhutdinov R. Multimodal Learning with Deep Boltzmann Machines // Journal of Machine Learning Research. 2014. no. 15. pp. 2949–2980.
12. Zurutuza U., Uribeetxeberria R. Intrusion Detection Alarm Correlation: A Survey // Proceedings of IADAT International Conference on Telecommunications and computer Networks. 2004. pp. 1–3.