© Gavrylenko S., Semenov S., Chelak V. Development of anomalous computer behavior detection method based on probabilistic automaton // Ukrainian Scientific Journal of Information Security, 2018, vol. 24, issue 3, pp. 163-168.

DOI: 10.18372/2225-5036.24.13427

РАЗРАБОТКА МЕТОДА ВЫЯВЛЕНИЯ АНОМАЛЬНОГО ПОВЕДЕНИЯ КОМПЬЮТЕРНОЙ СИСТЕМЫ НА ОСНОВЕ ВЕРОЯТНОСНОГО АВТОМАТА

Светлана Гавриленко, Сергей Семенов, Виктор Челак

Национальный технический университет «Харьковский политехнический институт»

ГАВРИЛЕНКО Светлана Юрьевна, к.т.н

Год и место рождения: 1963, г. Малин, Житомирской обл.
Образование: Харьковский политехнический институт.
Должность: профессор кафедры «Вычислительная техника и программирование» с 2013 года.
Научные интересы: защита информации в компьютерных системах и сетях.
Публикации: больше 100 научных трудов, среди которых, статьи, монографии, материалы симпозиумов и конференций.
E-mail: gavrilenko08@gmail.com

СЕМЕНОВ Сергей Геннадиевич, д.т.н., с.н.с.

Год и место рождения: 1972, г. Джанкой, Крымская обл.
Образование: Харьковский военный университет.
Должность: заведующий кафедры «Вычислительная техника и программирование» с 2013 года.
Научные интересы: защита информации в компьютерных системах и сетях.
Публикации: больше 100 научных трудов, среди которых, статьи, монографии, материалы симпозиумов и конференций.
E-mail: s_semenov@ukr.net

ЧЕЛАК Виктор Владимирович

Год и место рождения: 1996, г. Харьков, Харьковская обл.
Образование: С 2013 года студент Национального технического университета «Харьковский политехнический институт».
Научные интересы: антивирусные системы, реверс-инжиниринг, защита информации в компьютерных системах и сетях.
Публикации: более 20 научных трудов, среди которых статьи, материалы симпозиумов и конференций.
E-mail: victor.chelak@gmail.com

Аннотация. *В работе разработан метод выявления аномального поведения компьютерной системы на основе вероятностного автомата. Отличительной особенностью метода является адаптация процедуры генерации структуры автомата к ситуациям обнаружения однотипных сценариев, путем перестройки структуры автомата при обнаружении совпадений и пересчета вероятности переходов из состояния в состояние. Основными составляющими метода являются модель генерации структуры автомата и процедура его модификации. Входными данными автомата являются множество дискретных событий (системных вызовов, идентификаторов процессов или инструкций секций кода), присущих для определенного типа анормальности работы компьютерной системы и сгруппированные по классам. Первоначально генерируется структура автомата для одного из экземпляров классов, а затем эта структура перестраивается при анализе последующих экземпляров. Возможность перехода из состояния в состояние зависит от входного состояния и от значения вероятности перехода. Сгенерированная структура автомата в дальнейшем используется для выявления аномального поведения компьютерной системы. При возникновении аномалий с другими сценариями, структура автомата также может обновляться. Предложенный метод позволяет ускорит процесс выявления аномального поведения компьютера, а также обнаруживать аномалии компьютерной системы, профили сценариев которых лишь частично совпадают с экземплярами, используемыми при генерации структуры автомата. Полученные результаты исследований позволяют сделать вывод о возможности использования разработанного метода как дополнительного способа в эвристических анализаторах систем обнаружения аномалий.*

Ключевые слова: *вероятностный автомат, аномальное поведение компьютерной системы, эвристический анализатор, системы обнаружения аномалий.*

Вступление

Согласно полугодового отчета Cisco по информационной безопасности в первой половине 2018 рост количества и разновидности атак с применением компьютерных вирусов привело к колоссальным (53% атак нанесли ущерб на сумму более 500 тыс. долларов, 8% нанесли – на сумму более 5 млн. долларов) экономическим и имиджевым потерям в различных организациях во всем мире. Количество новых программных вирусов постоянно растет, несмотря на принятые во многих странах законы по борьбе с компьютерными преступлениями и разработку специальных программных средств защиты компьютерных систем. Поэтому разработка технологий противостояния компьютерным вирусам и методов фиксации аномального состояния компьютерной системы является **актуальной задачей**.

Повышение безопасности компьютерной системы (КС) представляет собой комплекс сложных и, как правило, взаимоувязанных научно-технических задач, решаемых на различных уровнях с использованием соответствующих организационно-технических мероприятий как по обеспечению целостности, доступности и конфиденциальности защищаемой информации, так и по обеспечению защиты применяемых методов сетевого управления, технологий доступа к предоставляемым сервисам и услугам, установленного порядка хранения и передачи данных.

Анализ литературы показал [1-6], что для решения указанных задач в настоящее время применяются различные механизмы. В основе их функционирования лежит сбор, анализ и обработка информации о событиях, связанных с безопасностью защищаемой КС, накопление полученных данных и принятие решения о состоянии системы с выявлением и возможным противодействием несанкционированному использованию вычислительных ресурсов на основе результатов проведенного анализа.

Основная сложность в использовании этих механизмов заключается в корректном выборе множества признаков, позволяющих разделить классы аномального поведения между собой и отделить их от нормального поведения. Механизмы, ориентированные на поиск искажений эталонной профильной информации, нечувствительны к последовательности сходных событий [2]. Все они имеют большие временные затраты на обучение и расчет коэффициентов, а также характеризуются высокой вероятностью ложного срабатывания.

Данные факторы дают основание разработчикам в реализации комплексных подходов выявления аномального поведения компьютерных систем. При этом наряду с интеллектуальными методами обработки данных (нейронные сети, нечеткая логика и др.) в оценке аномального состояния могут использоваться основные положения и подходы известной и хорошо себя зарекомендовавшей теории вероятностных автоматов [8-11]. Это, несмотря на повысившуюся сложность системы, позволит повысить точность принятия результирующего решения о состоянии компьютерной системы.

Таким образом, целью статьи является разработка метода выявления аномального поведения КС на основе теории вероятностных автоматов.

Разработка метода обнаружения аномалий

Предложенный в работе подход анализа поведения КС имеет два этапа. Первый этап заключается в генерировании структуры автомата и построении профилей аномальности поведения компьютерной системы. Второй этап – это проверка системы на наличие аномальности согласно сгенерированным на первом этапе профилям.

Генерирование структуры автомата (рис. 1) происходит за счет анализа входных данных согласно настройкам системы принятия решений (СПР).

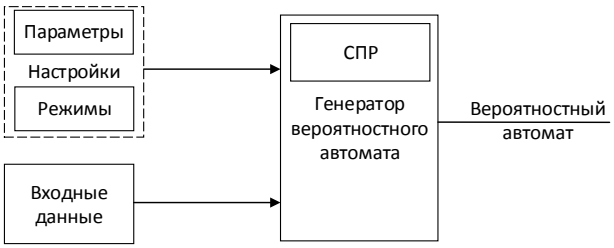

Рис. 1. Генерирование структуры автомата

СПР включает блок настроек и блок правил формирования структуры вероятностного автомата. В блоке настроек задаются различные режимы обработки входных данных. К режимам относятся: обработка данных (с учетом значений переменных или только инструкций), обработка стандартных функций (с параметрами или без) и др. В блоке параметров задаются: начальное значение вероятности переходов из состояния в состояние, размер входных данных в байтах, максимально допустимое количество итераций на один цикл и др. Наличие различных режимов позволяет преобразовать сценарии, содержащие циклы, ветвления, функции, в дискретную линейную последовательность событий.

Блок правил формирования структуры вероятностного автомата содержит правила формирования классов (конечных состояний) вероятностного автомата, правила оптимизации вероятностного автомата и др.

Следует заметить, что расширение учитываемых и математически формализуемых правил вероятностного автомата представляется довольно сложным в понимании процессом. Кроме того, исследования показали возможность формализации данных правил в виде графовых моделей. Поэтому в дальнейшем, при описании функционирования вероятностного автомата воспользуемся данным подходом.

Входными данными автомата являются множество дискретных событий (сценариев): системных вызовов, идентификаторов процессов или инструкций секций кода, присущих определенному типа аномальности работы компьютерной системы. Входные данные (экземпляры) группируются по классам, например, инструкции секций кода, присущие классу вредоносного программного обеспечения типа Trojan.

Первоначально для одного из экземпляров класса генерируется структура автомата. При этом на вход системы подается набор дискретных событий этого экземпляра класса. Каждое событие экземпляра сопоставляется с состоянием автомата s_i. Возможность перехода из состояния s_k в состояние s_m в момент времени t определяется входным состоянием x_k и значением коэффициента (маркера) K_m, задающего изменение («штраф» и «не штраф») вероятности перехода p_{km} из состояния s_k в состояние s_m. Изначально значения маркеров K и вероятностей переходов p задаются в блоке настроек автомата: K = 0, p = 1.

При анализе последующих событий этого класса, структура автомата дополняется новыми состояниями, которые соответствуют событиям последующих сценариев. При этом, на каждом шаге входные данные сопоставляются с уже сгенерированными условиями переходов автомата, и в случае обнаружения однотипных событий для разных сценариев, структура автомата перестраивается. Такой подход позволяет уменьшить количество состояний автомата, за счет объединения общих участков схемы переходов автомата.

Для примера, правила оптимизации вероятностного автомата математически можно представить в следующем виде:

$$P_{ij}(X,t)=\begin{cases}0, \ \textit{если}\ S_{ij}^{A}(X)=S_{Z}^{A},\\ P_{ij}(X,t-1), \ \textit{если}\ \left(S_{ij}^{A}(X)\neq S_{Z}^{A}\right)\ \&\&\ \left(S_{i,j}^{H}=S_{Z}^{H}\right),\\ P_{ij}(x,t-1)+K, \ \textit{если}\ S_{ij}^{H}\neq S_{Z}^{H}.\end{cases}$$

где $P_{ij}(X,t)$ – рассчитываемое значение вероятности переходов в таблице для i-го столбца и j-ой строки, $P_{ij}(X,t)$ –текущее значение таблицы вероятности, S_z^A все элементы множества состояний-предков, исключая элемент, с которым сравнивается данное множество, $S_{ij}^{A}(X)$ – элемент из множества предков, которому соответствует вероятность P_i, S_Z^H - множество наследников, включает все элементы кроме текущего $S_{ij}^{H}(X)$. K – маркерное значение (принимает значения -1 до 1).

Наличии однотипных событий в рамках разных экземпляров одного класса, описываемых одинаковой цепочкой переходов из состояния s_k в s_m состояние, приводит к увеличению маркера K_m, что в сою очередь увеличивает значение вероятности перехода p_{km} – процент совпадения входных данных с уже сгенерированными условиями переходов.

Процесс выявления однотипных сценариев регулируется переменной Rk (глубина рекурсии) и задается в блоке настройки СПР. Глубина рекурсии позволяет получить доступ на чтения Rk поколений предков и наследников, для определения идентичных участков сценариев.

На рис. 2 представлен пример содержимого памяти системы принятия решений для глубины рекурсии Rk = 2. (т.е. СПР хранит в себе информацию о соседях, пути к которым составляют от 1 до 2 переходов).

Пример генерирования структуры автомата для двух типов аномалий SK_1 и SK_2 и безопасного сценария SK_0, сценарии которых заданы в табл.1, приведен на рис. 3.

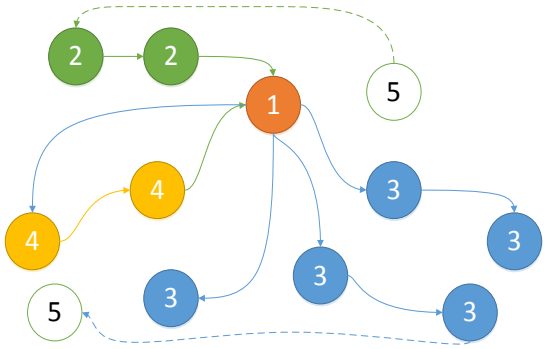

Рис. 2. Содержание памяти системы принятия решений для глубины рекурсии Rk = 2, где: 1 – указатель на текущее состояние автомата; 2 – состояния - предки для текущего состояния; 3 – состояния - наследники для текущего состояния; 4 – неопределенные состояния (к ним относиться состояния, которые система с заданным Rk, видит как в качестве предка, так и в качестве наследника): 5 - состояния автомата, которые не сохраняются в памяти СПР, но являются частью вероятностного автомата

Как видно из рис. 3, после оптимизации структуры автомата количество состояний, для заданных в табл.1 сценариев, уменьшилось до 58%. Значение маркера K, увеличилось для разных переходов с 0 до +4, +6. Увеличение значения маркера K повышает требование к проценту совпадения с входными данными с уже сгенерированными условиями переходов, что в свою очередь уменьшает вероятность ложного срабатывания.

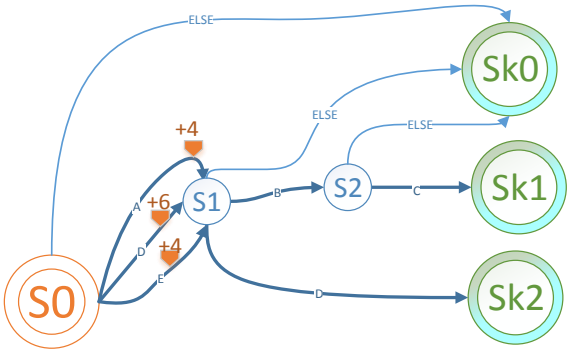

Рис. 3. Пример сгенерированной структуры автомата

Сгенерированная структура автомата в дальнейшем используется для выявления аномального поведения КС (рис.4).

Таблица 1

Примеры сценариев аномальности

Тип аномальности	Сценарии анормальностей для типов SK_1 и SK_2								
	Сценарий 1			Сценарий 2			Сценарий 3		
SK_1	A	B	C	D	B	C	E	B	C
SK_2	D	D	D	D	D	F	D	D	H

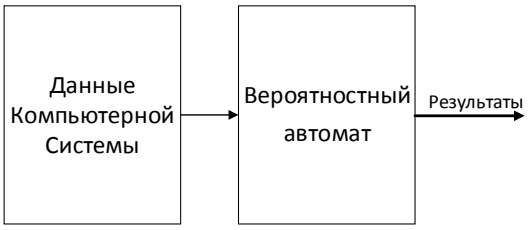

Рис. 4. Сгенерированная структура автомата для выявления аномального поведения КС

Результаты тестирование метода выявления аномального поведения КС

Для тестирования предложенного метода разработано программное обеспечение, позволяющее задавать различные параметры и режимы СПР автомата и сгенерировать структуру автомата для определенного класса аномалий. На рис 5. приведена экранная форма задания различных параметров и режимов СПР автомата.

Результаты проведенного эксперимента показали, что программа обнаружила аномалии поведения КС, сценарии которых полностью совпадали с разными вариантами сценариев классов SK_1 и SK_2 и сценарием нормального поведения SK_0. Кроме того, был обнаружен сценарий, частично совпадающий со сценарием класса SK_2 – SK_2_PartialMatch.

Результаты моделирования подтвердили возможность идентификации аномалий компьютерной системы, в том числе, профили которых лишь частично совпадают со сценариями, используемыми для генерирования структуры автомата. На рис. 6 приведены обобщенные результаты эксперимента. Для каждого из 500 сценариев, фиксировалось его время обнаружения. Процесс распознавания длился 500 с. Как видно из графика на 60 сек. было обнаружено 55 % сценариев, на 200 сек. – 78% сценариев, на 500 сек. – около 96,5% сценариев, приводящих к аномальному поведению КС.

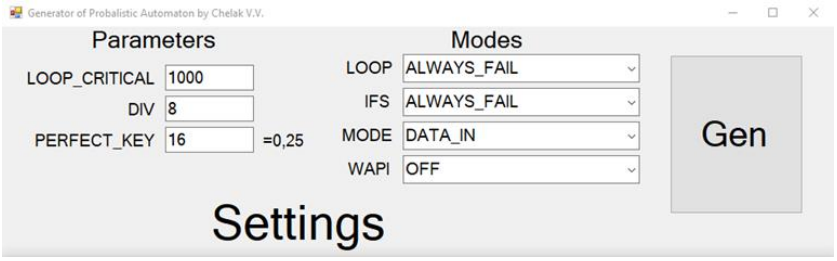

Рис. 5. Экранная форма задания различных параметров и режимов работы СПР автомата

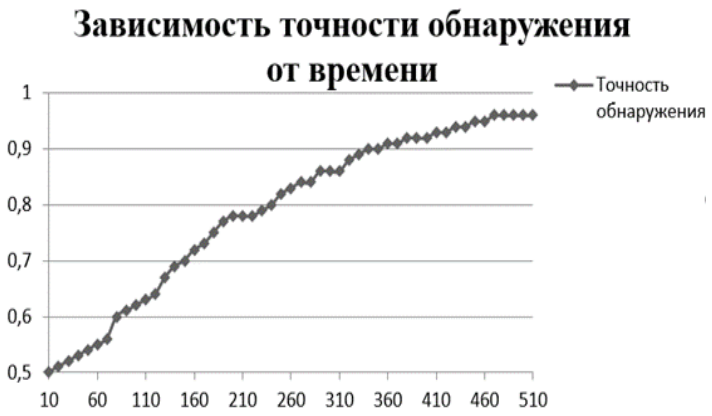

Рис. 6. Результаты тестирования системы

Выводы

В работе разработан метод фиксации аномального поведения КС на основе вероятностного автомата. Научной новизной метода является адаптация процедуры генерации структуры автомата к ситуациям обнаружения однотипных сценариев, путем перестройки структуры автомата при обнаружении совпадений и пересчета вероятности переходов из состояния в состояние. Основными составляющими метода являются модель генерации структуры автомата и процедура его модификации. Входными данными автомата являются множество дискретных событий (системных вызовов, идентификаторов процессов или инструкций секций кода), присущих для определенного типа анормальности работы компьютерной системы и сгруппированные по классам, например, разные сценарии DoS атак.

Первоначально генерируется структура автомата для одного из экземпляров классов, а затем эта структура перестраивается при анализе последующих экземпляров. Возможность перехода из состояния в состояние зависит от входного состояния и от значения вероятности перехода. Изначально, вероятности переходов задаются в настройках системы принятия решений, а затем каждый раз пересчитываются при обнаружении однотипных сценариев. Для обнаружения однотипных сценариев, задается значение области видимости граф-схемы автомата Rk. Исходя из этого значения, на каждом шаге, относительно текущего состояния автомата, просматри-

ваются Rk состояний предков и наследников граф-схемы автомата и при обнаружении совпадений структура автомата перестраивается. Такой подход позволяет уменьшить количество состояний автомата, так как однотипные сценарии разных экземпляров описываются одним и тем же участком граф-схемы переходов автомата.

Сгенерированная структура автомата в дальнейшем используется для выявления аномального поведения КС. При возникновении аномалий с другими сценариями, структура автомата также может обновляться.

Результаты тестирования показали, что предложенный метод позволяет ускорит процесс выявления аномального поведения компьютера и обнаруживать аномалии компьютерной системы, в том числе, профили которых лишь частично совпадают с экземплярами, используемыми при генерации структуры автомата. Предложенный метод может быть использован как дополнительный способ в общей системе обнаружения аномального поведения КС.

Литература

[1]. О. Шелухин, Д. Сакалема, А. Филинова, *Обнаружение вторжений в компьютерные сети (сетевые аномалии): учебное пособие для вузов*, М.: Горячая линия-Телеком, 2013, 220 с.

[2]. Т. Шипова, В. Босько, И. Березюк, Ю. Пархоменко, "Анализ современных методов обнаружения вторжений в компьютерные системы", *Системи обробки інформації: зб. наук. пр.*, Харьков: ХУ ПС, Вип. 1 (139), С. 133-137, 2016.

[3]. S. Zacher, P. Ryba, "Anomaly detection in server metrics with use of one-sided median algorithm", *JACSM*, vol. 9, no. 1, pp. 5-22, 2017.

[4]. L. Akoglu, H. Tong, D. Koutra, "Graph based anomaly detection and description: a survey", *Data Mining and Knowledge Discovery*, vol. 29(3), pp. 626-688, 2015.

[5]. H. Al-Hamami, G. Al-Saadoon, "Development of a network-based: Intrusion Prevention System using a Data Mining approach", *Science and Information Conference, London*, pp. 641-644, 2013.

[6]. C. Kruegel, D. Mutz, F. Valeur, G. Vigna, "On the detection of anomalous system call arguments", *in In Proc. of the 8th European Symposium on Research in Computer Security*. Springer-Verlag. pp. 326-343, 2003.

[7]. М. Рабин, "Вероятностные автоматы", *Кибернетический сборник*, Вып. 9, М.: Иностранная литература, С. 123-141, 1964.

[8]. A. Maier, O. Niggemann, R. Just, M. J¨ager, *Anomaly Detection in Production Plants using Timed Automata*. [Electronic resource]. Online: https://www. researchgate.net /publication/257365001_Anomaly_Detection_in_Production_Plants_using_Timed_Automata.

[9]. F. Kepler, S. Mergen, C. Billa, "Simple variable length n-grams for probabilistic automata learning. Journal of Machine Learning Research", *Workshop and Conference Proceedings, ICGI'12*, pp. 254-258, 2012.

[10]. S. Verwer, M. Weerdt, C. Witteveen, "A likelihood-ratio test for identifying probabilistic deterministic real-time automata from positive data", *In Proceedings of ICGI'10, volume 6339 of LNCS, Springer-Verlag*, pp. 203-216, 2010.

[11]. Kui Xu, Danfeng Yao, Barbara Ryder, Ke Tian, *Probabilistic Program Modeling for High-Precision Anomaly Classification*. [Electronic resource]. Online: http:// people. cs. vt. edu/danfeng/papers/HMM-CSF-15-Yao.pdf.

УДК 004.732.056

Гавриленко С.Ю., Семенов С.Г., Челак В.В. Розробка методу виявлення аномальної поведінки комп'ютерної системи на основі імовірнісного автомата

Анотація. *В роботі запропоновано метод виявлення аномальної поведінки комп'ютерної системи на основі імовірнісного автомата. Основними складовими методу є модель генерації структури автомата і процедура його модифікації Відмінною особливістю методу є адаптація процедури генерації структури автомата до ситуацій виявлення однотипних сценаріїв, шляхом перебудови структури автомата при виявленні збігів і перерахунку ймовірності переходів зі стану в стан. Вхідними даними автомата є безліч дискретних подій (системних викликів, ідентифікаторів процесів або інструкцій секцій коду), властивих для певного типу аномалії роботи комп'ютерної системи і згруповані за класами. Спочатку генерується структура автомата для одного з примірників класів, а потім ця структура перебудовується при аналізі наступних примірників. Можливість переходу зі стану в стан залежить від вхідного стану і від значення ймовірності переходу. Згенерована структура автомата в подальшому використовується для виявлення аномальної поведінки комп'ютерної системи. При виникненні аномалій з іншими сценаріями, структура автомата також може оновлюватися. Запропонований метод дозволяє прискорить процес виявлення аномальної поведінки комп'ютера, а також виявляти аномалії комп'ютерної системи, профілі сценаріїв яких лише частково збігаються з екземплярами, використовуваними при генерації структури автомата. Отримані результати досліджень дозволяють зробити висновок про можливість використання розробленого методу як додаткового способу в евристичних аналізаторах систем виявлення аномалій.*

Ключові слова: *імовірнісний автомат, аномальна поведінка комп'ютерної системи, евристичний аналізатор, системи виявлення аномалій.*

Gavrylenko S., Semenov S., Chelak V. Development of anomalous computer behavior detection method based on probabilistic automaton

Abstract*. The paper proposes a method for identifying the anomalous behavior of a computer system based on probabilistic automaton. The main components of the method are the model of generation of the structure of the automaton and its modification procedure. The defining feature of the method is adaptation of automaton structure generation procedure for detecting scenarios of the same type, by restructuring the structure of the automaton upon a match and by recalculation of the state transition probabilities. Input data of the automaton consist of discrete events (system calls, process IDs or sections of code instructions), typical for a certain*

class of anomalous behavior, and grouped by type. The automaton structure is first created in accordance with one of the instances of a class, and then restructured during the analysis of other instances. Possibility of state transition depends on the input state and transition probability value. Generated automaton structure is used to detect anomalous computer system behavior. Automaton structure can be updated, if an anomaly occurs with different scenarios. Proposed method allows to speed up detecting anomalous computer behavior, as well as to detect computer system anomalies, scenario profiles of which only partially match with instances used for generation the structure of the automaton. Obtained research results allow us to conclude about the possibility of using this method in heuristic analyzers of anomaly detection systems.
Keywords: *probabilistic automaton, anomalous computer system behavior, heuristic analyzer, anomaly detection system.*

Отримано 16 *жовтня 2018 року*, затверджено редколегією *30 листопада 2018 року*