ТЕХНИКАЛЫҚ ҒЫЛЫМДАР
ТЕХНИЧЕСКИЕ НАУКИ
TECHNICAL SCIENCE

CENTRAL ASIAN SCIENTIFIC JOURNAL
ISSN 3135-3061
2026. №4 (32), Том 2. С. 512–517
DOI: https://doi.org/10.5281/zenodo.24362138
Сайт журнала: https://cajournal.kz

УДК 004.056.53:004.94

РАЗРАБОТКА ПЛАТФОРМЫ МОДЕЛИРОВАНИЯ И АНАЛИЗА АТАК СОЦИАЛЬНОЙ ИНЖЕНЕРИИ

А.С. Есенов, Е. Мухтаров, А. Аукен, А. Байтурин
НАО «Казахский национальный исследовательский технический университет имени К. И. Сатпаева» (Satbayev University), г. Алматы, Республика Казахстан

Есенов А. С. — НАО «Казахский национальный исследовательский технический университет имени К. И. Сатпаева» (Satbayev University), г. Алматы, Республика Казахстан.

Мухтаров Е. — НАО «Казахский национальный исследовательский технический университет имени К. И. Сатпаева» (Satbayev University), г. Алматы, Республика Казахстан.

Аукен А. — НАО «Казахский национальный исследовательский технический университет имени К. И. Сатпаева» (Satbayev University), г. Алматы, Республика Казахстан.

Байтурин А. — НАО «Казахский национальный исследовательский технический университет имени К. И. Сатпаева» (Satbayev University), г. Алматы, Республика Казахстан.

© А.С. Есенов, Е. Мухтаров, А. Аукен, А. Байтурин, 2026

Аннотация: В настоящем исследовании представлена архитектурно-методологическая концепция специализированной программной платформы, предназначенной для автоматизированного конструирования, имитационного воспроизведения и многофакторной аналитической оценки социотехнических векторов компрометации корпоративных информационных систем. Актуальность работы обусловлена устойчивым доминированием атак, эксплуатирующих когнитивно-поведенческие уязвимости персонала, над классическими эксплойтами программно-аппаратного уровня. В статье сформулирована математическая модель квантификации человеческого фактора на основе динамического индекса уязвимости (Cognitive Vulnerability Score), учитывающего темпоральные характеристики отклика сотрудника, ретроспективный профиль инцидентов и индекс контекстной сложности атакующего сценария. Предложена микросервисная архитектура на базе языка Python с интеграцией механизмов криптографической верификации журналируемых событий (HMAC-SHA256) и ролевого разграничения полномочий в парадигме Zero Trust. Проведен сравнительный аудит разработанного прототипа относительно коммерческих и открытых аналогов (Gophish, KnowBe4, Hoxhunt), подтвердивший преимущества локализованного развертывания в контуре безопасности предприятия.

Ключевые слова: социальная инженерия, имитационное моделирование атак, социотехнический аудит, поведенческая аналитика, человеческий фактор, фишинг, информационная безопасность, Zero Trust.

Аңдатпа. Мақалада корпоративтік ақпараттық жүйелерге бағытталған әлеуметтік инженерия шабуылдарын модельдеуге және талдауға арналған бағдарламалық платформаның сәулеті мен әдіснамасы ұсынылған. Қызметкерлердің когнитивті осалдық индексін бағалаудың математикалық моделі мен Python негізіндегі микросервистік құрылым сипатталған.

Abstract. This paper presents the architectural design and mathematical modeling methodology of an automated software platform for simulating and analyzing social engineering

attacks. A composite Cognitive Vulnerability Score (CVS) metric and a Python-based microservice infrastructure with HMAC-verified audit logging are proposed to enhance organizational resilience against human-centric cyber threats.

1. Введение и постановка проблемы

Современный ландшафт киберугроз характеризуется фундаментальной трансформацией векторов первичного проникновения в защищенные корпоративные инфраструктуры. По мере совершенствования механизмов эшелонированной технической защиты — включая системы обнаружения и предотвращения вторжений нового поколения (NGFW/IDS), средства поведенческого анализа конечных точек (EDR/XDR) и криптографические протоколы аутентификации — вычислительная и финансовая стоимость эксплуатации программно-аппаратных уязвимостей нулевого дня (0-day) для злоумышленников существенно возросла. Вследствие данной асимметрии доминирующим каналом компрометации периметра безопасности стала целенаправленная эксплуатация когнитивных, психологических и организационных искажений персонала, объединяемая термином «социальная инженерия».

В отличие от детерминированных программных сбоев, социотехнические уязвимости обладают стохастической природой и напрямую зависят от психоэмоционального состояния оператора, уровня корпоративной культуры безопасности и контекстной релевантности атакующего воздействия. Классические средства периметральной фильтрации трафика демонстрируют ограниченную результативность при противодействии полиморфным атакам целевого фишинга (Spear Phishing), компрометации деловой переписки (Business Email Compromise, BEC) и гибридным многоканальным воздействиям (сочетание голосового вишинга, мессенджер-манипуляций и поддельных порталов единого входа SSO).

Таким образом, перед службами информационной безопасности (SOC и подразделениями комплаенс-аудита) встает научно-практическая задача перехода от эпизодического инструктажа сотрудников к непрерывному инструментальному измерению и проактивному моделированию устойчивости персонала. Целью настоящей работы является проектирование и программная реализация автономной платформы моделирования и анализа атак социальной инженерии, обеспечивающей замкнутый цикл «имитация угрозы — телеметрия реакции — квантификация риска — адаптивная коррекция поведения» с полным соблюдением требований локализации персональных данных и методических нормативов СТ КазНИТУ 09-2023.

2. Критический анализ существующих платформ и предпосылки разработки

Для обоснования архитектурных решений проектируемой системы был выполнен многокритериальный сравнительный аудит представленных на мировом рынке программных комплексов социотехнического тестирования: открытого фреймворка Gophish, а также проприетарных корпоративных облачных платформ KnowBe4 и Hoxhunt.

Анализ функциональных возможностей открытой системы Gophish показал, что при высокой прозрачности исходного кода и возможности развертывания в изолированном контуре (On-Premise), данный инструмент предоставляет лишь базовый транспортный механизм рассылки электронных писем и фиксации бинарных событий (открытие письма, переход по гиперссылке, отправка POST-запроса). В Gophish отсутствует подсистема многофакторного скоринга уязвимости подразделений, не реализованы механизмы автоматического назначения компенсационных обучающих модулей в момент совершения ошибки, а также не предусмотрена криптографическая защита журналов аудита от несанкционированной модификации.

В свою очередь, коммерческие решения класса Enterprise (KnowBe4, Hoxhunt) обладают развитыми средствами геймификации и обширными библиотеками шаблонов, однако характеризуются рядом критических ограничений для отечественных предприятий и академических учреждений Республики Казахстан:

- Экстерриториальная обработка телеметрии: функционирование исключительно по модели SaaS в зарубежных центрах обработки данных создает прямые

комплаенс-риски нарушения суверенитета данных и раскрытия организационно-штатной структуры предприятия третьим сторонам.

- Отсутствие региональной и лингвистической адаптации: шаблоны социотехнических приманок не учитывают специфику локального документооборота, двуязычной деловой среды (казахский и русский языки) и региональных сервисов цифровой идентификации.
- Закрытость алгоритмов оценки риска: проприетарные модели расчета индекса уязвимости («чёрный ящик») не позволяют интегрировать специфические весовые коэффициенты критичности активов конкретного предприятия или кафедры.

В таблице 1 представлена сводная матрица сопоставления функционально-технических характеристик существующих решений и разрабатываемой в Сатбаев Университете платформы.

Таблица 1 — Сравнительный анализ платформ имитации и анализа атак социальной инженерии

Критерий оценки	Gophish (Open-Source)	KnowBe4 (Enterprise SaaS)	Hoxhunt (AI SaaS)	Разрабатываемая платформа
Модель развертывания	On-Premise (локально)	Публичное облако (SaaS)	Публичное облако (SaaS)	**On-Premise / Гибридный контур**
Многофакторный скоринг риска	Отсутствует (только сырые метрики)	Статический Risk Score	Динамический поведенческий скоринг	**Адаптивный индекс CVS с весами активов**
Локализация (KZ/RU контекст)	Ручная верстка шаблонов	Частичная (автоперевод)	Ограниченная	**Полная нативная поддержка KZ/RU сценариев**
Защита логов аудита	Стандартная СУБД (SQLite/MySQL)	Внутренний облачный лог	Внутренний облачный лог	**Криптографическая цепочка HMAC-SHA256**
Модель разграничения доступа	Базовая (Admin/User)	RBAC	RBAC	**Гибридная RBAC + ABAC (Zero Trust)**

3. Архитектура разрабатываемой платформы и механизмы безопасности

Проектируемая платформа построена на принципах слабосвязанной микросервисной архитектуры с использованием высокопроизводительного стека языка Python. Выбор технологического базиса обусловлен богатой экосистемой библиотек статистической обработки данных, криптографических примитивов и асинхронных сетевых фреймворков. Структурно система разделена на пять изолированных функциональных контуров, взаимодействующих через защищенный шлюз REST API.

1. Подсистема оркестрации и синтеза сценариев (Scenario Synthesis & Campaign Engine). Обеспечивает конструирование многошаговых социотехнических кампаний, имитирующих различные психологические триггеры: административный авторитет, дефицит времени, финансовую срочность и техническую поддержку. Модуль поддерживает параметрическую генерацию фишинговых писем, целевых лендинговых страниц перехвата учетных данных и имитацию вложений с макрос-триггерами, безопасными для реальной инфраструктуры, но фиксирующими факт запуска.

2. Транспортно-телеметрический узел (Telemetry & Interception Collector). Реализует асинхронную доставку имитационных пакетов через выделенный SMTP-релей и перехват обратных обращений тестируемых пользователей. Узел регистрирует детализированные временные метки (timestamps) с точностью до миллисекунды, параметры пользовательского окружения (User-Agent, отпечаток браузера) и глубину взаимодействия с фальшивой формой аутентификации без сохранения реальных паролей в открытом виде.

3. Ядро поведенческой аналитики и скоринга (Behavioral Risk Analytics Core). Выполняет агрегацию первичной телеметрии, корреляцию действий сотрудника с его

должностным профилем и пересчет интегральных показателей уязвимости структурных подразделений в режиме реального времени.

4. Модуль мгновенной адаптивной коррекции (Just-in-Time Micro-Learning). Реализует концепцию контекстного обучения в момент ошибки. При фиксации критического действия (например, отправка учетных данных на имитационном портале) пользователь автоматически перенаправляется на интерактивную страницу разбора, где визуализируются конкретные индикаторы компрометации (IoC), которые были проигнорированы в данном письме или доменном имени.

5. Подсистема безопасности, контроля доступа и неизменяемого аудита (Zero Trust & Audit Core). Поскольку платформа социотехнического тестирования аккумулирует чувствительные сведения об уязвимостях конкретного персонала, внутренний контур самой системы спроектирован в соответствии с парадигмой нулевого доверия (Zero Trust). Аутентификация операторов осуществляется посредством короткоживущих токенов JWT, а авторизация объединяет ролевую (RBAC) и атрибутивную (ABAC) модели доступа.

Для гарантирования юридической и доказательной целостности результатов тестирования в платформу интегрирован специализированный механизм криптографического журналирования операций базы данных. Каждая запись о результате симуляции связывается с предыдущей записью посредством вычисления кода аутентификации сообщения HMAC на базе хеш-функции SHA-256 и динамической криптографической соли (генерируемой через системный источник энтропии os.urandom), что исключает возможность ретроспективной фальсификации отчетов аудита.

4. Математическая модель оценки когнитивной уязвимости персонала

В отличие от простейших бинарных метрик вида «процент кликнувших пользователей» (Click-Through Rate, CTR), которые не отражают реальную тяжесть последствий для бизнеса, в разрабатываемой платформе предложен композитный индекс когнитивной уязвимости сотрудника — Cognitive Vulnerability Score (CVS).

Индивидуальная вероятность успешной компрометации сотрудника i при реализации сценария социальной инженерии j формализуется следующим аналитическим выражением:

$$CVS(i, j) = A_crit(i) \times [\, w_1 \cdot D_act(i, j) + w_2 \cdot exp(-\lambda \cdot \Delta t(i, j)) + w_3 \cdot H_rec(i) \,] \times (1 / S_comp(j))$$

где переменные и параметры модели имеют следующую физическую и информационную интерпретацию:

- A_crit(i) ∈ [0.5; 2.0] — коэффициент привилегированности и критичности информационных активов, к которым имеет доступ сотрудник i (определяется автоматически на основе членства в группах безопасности каталога Active Directory / LDAP);
- D_act(i, j) ∈ [0; 1] — дискретный коэффициент глубины компрометации в рамках сценария j (0.0 — игнорирование письма; 0.2 — открытие сообщения; 0.6 — переход по фишинговой гиперссылке; 1.0 — ввод аутентификационных данных или запуск исполняемого макроса; -0.3 — проактивная отправка уведомления о подозрительном письме в службу ИБ через кнопку Phish-Report);
- Δt(i, j) — латентность реакции (временной интервал между доставкой стимула и совершением целевого ошибочного действия). Экспоненциальный член exp(-λ·Δt) отражает импульсивность реакции: мгновенный переход по ссылке в первые секунды свидетельствует о полном отсутствии критической верификации заголовков и URL;
- H_rec(i) ∈ [0; 1] — коэффициент рецидивности, рассчитываемый методом экспоненциально взвешенного скользящего среднего (EWMA) по результатам предыдущих N имитационных кампаний;
- S_comp(j) ∈ [1; 5] — индекс когнитивной сложности сценария j (обратно пропорционален заметности индикаторов фишинга: чем изощреннее атака, тем меньший штрафной вес начисляется сотруднику при ошибке);
- w_1, w_2, w_3 — нормированные весовые коэффициенты значимости факторов ($w_1 + w_2 + w_3 = 1$).

Агрегированный показатель защищенности структурного подразделения или университета в целом определяется как средневзвешенное значение индивидуальных индексов CVS с учетом топологии внутренних коммуникаций. Такой математический аппарат позволяет офицерам информационной безопасности выявлять не только отдельных уязвимых операторов, но и системные «узлы риска» (например, бухгалтерию, отдел кадров или деканат), обладающие максимальным произведением когнитивной уязвимости на уровень привилегий доступа.

5. Экспериментальная апробация и обсуждение результатов

Для верификации предложенной архитектуры и математической модели скоринга было проведено пилотное моделирование на тестовой выборке из 120 пользователей, разделенных на три профильные когорты в условиях изолированного академического полигона. Программа испытаний включала три последовательные волны имитационных кампаний с интервалом в 14 суток, различающихся по вектору психологического воздействия:

Волна 1 (Базовый срез — сценарий «Срочное обновление пароля корпоративной почты»): эксплуатация триггера страха потери доступа к учетной записи. На данном этапе модуль адаптивного микрообучения был отключен для фиксации естественного фонового уровня уязвимости. Доля переходов по ссылке составила 34,2%, доля ввода учетных данных — 19,1%, а средний индекс CVS по выборке достиг 0,68.

Волна 2 (Промежуточный срез — сценарий «Подписание приказа / ведомости во вложении»): после завершения первой волны был активирован модуль Just-in-Time Micro-Learning. Несмотря на повышение сложности сценария (использование контекстно-знакомых формулировок учебного процесса), доля критических ошибок (ввод данных / открытие макроса) снизилась до 9,8%, а время первичной верификации письма пользователем (Δt) увеличилось в среднем в 3,4 раза, что свидетельствует о формировании навыка визуальной инспекции адреса отправителя и доменной зоны.

Волна 3 (Контрольный срез — комбинированный целевой сценарий Spear Phishing): по итогам трех итераций интегральный показатель критической компрометации стабилизировался на отметке 4,1%, при этом доля проактивных сообщений в службу безопасности через плагин репортинга выросла с исходных 5,0% до 46,7%. Средний индекс когнитивной уязвимости CVS снизился с 0,68 до 0,21 (снижение интегрального социотехнического риска на 69,1%).

Таблица 2 — Динамика показателей социотехнической устойчивости в ходе экспериментальной апробации (N = 120)

Этап моделирования	Переход по ссылке (CTR), %	Ввод учетных данных (Compromise), %	Проактивный репортинг (Report Rate), %	Средний индекс уязвимости (CVS)
Волна 1 (Базовый аудит без обучения)	34,2 %	19,1 %	5,0 %	0,68
Волна 2 (После активации JIT-обучения)	17,5 %	9,8 %	24,2 %	0,39
Волна 3 (Контрольный целевой сценарий)	8,3 %	4,1 %	46,7 %	0,21

6. Заключение и перспективы развития проекта

В результате проведенного исследования разработана и апробирована комплексная программная платформа моделирования и анализа атак социальной инженерии, ориентированная на повышение киберустойчивости корпоративных и образовательных структур Республики Казахстан. Научная новизна и практическая значимость работы заключаются в следующем:

- Спроектирована автономная микросервисная архитектура на базе Python, обеспечивающая полный цикл социотехнического аудита в закрытом контуре организации без передачи чувствительной телеметрии во внешние облачные сервисы.
- Предложена оригинальная математическая модель расчета индекса когнитивной уязвимости (CVS), объединяющая глубину ошибки, латентность реакции оператора, историю инцидентов и критичность доступных информационных активов.
- Реализованы механизмы криптографической защиты журналов аудита на базе HMAC-SHA256 и разграничения доступа по модели Zero Trust (RBAC/ABAC), исключающие несанкционированный доступ к профилям уязвимости сотрудников.
- Экспериментально доказана эффективность модуля мгновенного контекстного микрообучения (Just-in-Time Learning), позволившего за три итерации снизить интегральный риск успешной социотехнической компрометации на 69,1%.

Дальнейшее развитие платформы предполагает интеграцию больших языковых моделей (LLM), развернутых локально, для автоматизированного синтеза контекстно-зависимых сценариев на основе открытых корпоративных данных (OSINT-профилирование), а также расширение каналов тестирования за счет модулей имитации атак через корпоративные мессенджеры и голосовые шлюзы.

Список использованных источников:

1. Стандарт организации СТ КазНИТУ 09-2023. Работы учебные и научно-исследовательские. Общие требования к построению, изложению и оформлению. — Алматы: Satbayev University, 2023. — 48 с.

2. Hadnagy C. Social Engineering: The Science of Human Hacking. — 2nd ed. — Indianapolis: John Wiley & Sons, 2018. — 320 p.

3. ISO/IEC 27001:2022. Information security, cybersecurity and privacy protection — Information security management systems — Requirements. — Geneva: ISO, 2022.

4. Rose S., Borchert O., Mitchell S., Connelly S. Zero Trust Architecture // NIST Special Publication 800-207. — Gaithersburg: National Institute of Standards and Technology, 2020. — 59 p.

5. Salahdine F., Kaabouch N. Social Engineering Attacks: A Survey // Future Internet. — 2019. — Vol. 11, No. 4. — Article 89. — P. 1–17.

6. Wright J., Cache J. Hacking Exposed Wireless: Wireless Security Secrets & Solutions. — 3rd ed. — New York: McGraw-Hill Education, 2015. — 544 p.

© Central Asian Scientific Journal, 2026