Секция «44.41 LegalTech. Цифровая криминалистика»

Формализация и оптимизация интегрированных систем защиты текстовой информации в условиях антагонистического взаимодействия

Научный руководитель – Погребинская Мария Николаевна

Погребинская Мария Николаевна
Студент (специалист)
Московский государственный технический университет имени Н.Э. Баумана, Москва, Россия
E-mail: mashapogrebinskaya@yandex.ru

OCR-технологии стали инфраструктурными для оцифровки документов и DLP-контроля: для печатного текста при качественных сканах указывается точность распознавания, превышающая 99% при условии последующей проверки человеком [4]. Практика внедрения OCR в DLP подчеркивает критическую роль условий съемки и предобработки изображения [9]. Одновременно индустрия усиливает извлечение текста из «сложных» изображений за счет нейросетевой локализации текстовых зон и алгоритмического восстановления структуры документа [8]. Предлагается формальная минимакс-постановка оптимизации интегрированных систем защиты текста как антагонистической игры «защитник–атакующий» с многокритериальной функцией риска.

В прикладной информационной безопасности защита информации трактуется как совокупность мер и средств противодействия угрозам и уязвимостям при ограниченных ресурсах [2]. Для юридически значимых текстов конфликт критериев проявляется как противоречие между читаемостью (легитимное ознакомление, воспроизводимость предъявления) и стойкостью к автоматизированному извлечению и/или смысловой подмене текста.

Объектом исследования выступают интегрированные системы защиты текста, сочетающие физические и цифровые механизмы. В прикладной архитектуре такие системы могут включать многоуровневую печать со специальными слоями и искажениями (например, AM/FM-полутоновые узоры, УФ/ИК-чернила, микродеформации контуров букв, поляризационные слои), а для электронных копий — DRM-контур чтения с запретом копирования, антискриншотом и динамическими водяными знаками. Задача оптимизации состоит в выборе конфигурации защиты, которая минимизирует риск извлечения при сохранении доказательственной пригодности документа.

Выбор конфигурации защиты предлагается описывать как антагонистическое взаимодействие «защитник–атакующий». Теория игр задает аппарат стратегий и смешанных распределений в условиях неопределенности [3]. В кибербезопасности показана применимость игровых моделей к оптимизации распределения ограниченных ресурсов защиты при адаптивных угрозах [1]. Методически близка также задача выбора комбинаций мер контроля под бюджетом и с учетом зависимостей между мерами: она формализуется как двухлицевая игра с нулевой суммой и отбором допустимых комбинаций [6]. Эти подходы переносимы на конфигурирование интегрированной защиты текста, где «меры контроля» соответствуют модулям защиты, а «профили атакующего» — классам сценариев извлечения.

Пусть S — множество допустимых конфигураций (комбинаций модулей защиты), A — множество атакующих сценариев извлечения (сканирование/пересъем, предобработка, OCR, пост-коррекция). Тогда оптимизация задается как минимизация наихудшего риска: $\min_{\{s\in S\}} \max_{\{a\in A\}} (R(s,a) + \lambda\cdot C(s) + \mu\cdot H(s))$, где R(s,a) — риск успешного извлечения и/или смысловой подмены, C(s) — стоимость внедрения и сопровождения, H(s) — штраф за ухудшение читаемости и «правовой пригодности» (возможность ознакомления и воспроизводимость

предъявления). Коэффициенты λ и μ задают нормативно-организационный компромисс между стойкостью и условиями легитимного использования.
Угрозмодель должна включать адаптивные атаки на OCR. Показано, что малые модификации изображений печатного текста могут приводить к целевому искажению распознанного содержания при визуальной сохранности для человека [7]. Описаны атаки, маскирующие возмущения под водяные знаки и демонстрирующие высокую успешность при «естественном» внешнем виде результата [5]. Следовательно, R(s,a) целесообразно задавать многокомпонентно и включать: (а) вероятность извлечения текста в заданной атакующей цепочке; (б) вероятность смысловой подмены/искажения в выходном тексте (в т.ч. на уровне ключевых юридических реквизитов); (в) устойчивость к предобработке, характерной для DLP-контуров и практических пайплайнов OCR [9]; (г) оценку деградации читаемости для человека, поскольку «полная нераспознаваемость машиной» не должна достигаться ценой утраты юридически значимой доступности.
Минимакс-постановка «защитник–атакующий» переводит проектирование интегрированных систем защиты текста из эвристического выбора модулей в воспроизводимую процедуру оптимизации по критериям риска, стоимости и читаемости, релевантную задачам LegalTech и цифровой криминалистики [2].

Источники и литература

1) Афанасьева М.В., Наследов С.Е., Русецкас В.С.. Теория игр для поддержки решений в SOC // Защита информации. INSIDE. 2025. № 6. С. 30–37.
2) Варфоломеев А.А.. Основы информационной безопасности: учебное пособие. М.: Российский университет дружбы народов, 2008. 412 с.
3) Диксит А., Скит С., Рейли-мл. Д.. Стратегические игры. Доступный учебник по теории игр / пер. с англ. Н. Яцюк. М.: Манн, Иванов и Фербер, 2017. 880 с.
4) Ткаченко В.В., Тарлычева П.Н.. Информационная система автоматизированного распознавания реквизитов экономических документов на основе OCR-алгоритмов // Вестник Академии знаний. 2023. № 54(1). С. 239–244.
5) Chen L., Sun J., Xu W.. FAWA: Fast Adversarial Watermark Attack on Optical Character Recognition (OCR) Systems. 2020. arXiv:2012.08096.
6) Léveillé D., Jaskolka J.. A Game-Theoretic Approach for Security Control Selection // EPTCS. 2024. Vol. 409. P. 103–119. doi:10.4204/EPTCS.409.11.
7) Song C., Shmatikov V.. Fooling OCR Systems with Adversarial Text Images. 2018. arXiv:1802.05385.
8) CNews. ГК InfoWatch запатентовала технологию распознавания текста на сложных изображениях. [Электронный ресурс] URL: https://www.cnews.ru/news/line/2026-02-04_gk_infowatch_zapatentovala_tehnologiyu. Дата обращения: 04.02.2026.
9) SolarSecurity. Сложности применения технологий OCR в DLP-системах, или Как мы OCR готовим // Хабр. [Электронный ресурс] URL: https://habr.com/ru/companies/solarsecurity/articles/460881/ (дата обращения: 04.02.2026).