DOI: 10.37791/2687-0649-2025-20-1-125-139

Модель стеганографического встраивания в файлы с иерархической структурой

С.В. Белим[1*]***, С.Н. Мунько***[1]***, С.Ю. Белим***[1]
[1]*Омский государственный технический университет, Омск, Россия*
[*]*sbelim@mail.ru*

Аннотация. Стеганографические методы всегда ориентированы на конкретный формат файла, используемого в качестве контейнера. Наибольшие трудности возникают при встраивании в текстовые документы с разметкой. В статье предложена модель встраивания в управляющие теги структурированных текстовых документов. Модель использует древовидную структуру документа и осуществляет встраивание в свободные листовые вершины. Такой подход позволяет добавлять скрытые данные, не влияющие на штатное отображение документа. На основе данной модели предложено два стеганографических метода. Первый метод встраивает скрытые данные в теги html-документа, добавляя неотображаемые теги и стилевые классы к свободным листовым вершинам. Для обнаружения встроенных данных используется идентификатор встраивания, роль которого играет имя нового класса. Для защиты от стегоанализа имена новых классов создаются с использованием ключа встраивания и стойкой хеш-функции. Формат идентификаторов выбирается таким образом, чтобы совпадать с форматом имен исходного документа. Такой подход формирования имен позволяет распределять блоки скрытого сообщения случайным образом по свободным листовым вершинам. Второй метод предназначен для стеганографического встраивания в xml-документы. Скрытые данные добавляются в атрибуты свободных листовых вершин. Для работы метода требуется два новых атрибута. Оба атрибута описываются в виде дополнительной структуры, не отличимой от присутствующих в документе. Идентификатор встраивания также формируется с помощью ключа встраивания и номера встроенного блока. Для представления данных используется алгоритм шифрования, что требует введения дополнительного ключа. Оба метода используют маскировку встроенных данных для противодействия стегоанализу исходного кода. Стегоанализ таких методов имеет экспоненциальную алгоритмическую сложность, поэтому оба метода применимы только к большим файлам.

Ключевые слова: стеганография, модель встраивания, иерархический файл, xml-формат, html-формат

Для цитирования: *Белим С.В., Мунько С.Н., Белим С.Ю.* Модель стеганографического встраивания в файлы с иерархической структурой // Прикладная информатика. 2025. Т. 20. № 1. С. 125–139. DOI: 10.37791/2687-0649-2025-20-1-125-139

Steganographic embedding model in files with hierarchical structure

S.V. Belim[1*]***, S. Munko***[1]***, S.Yu. Belim***[1]
[1]*Omsk State Technical University, Omsk, Russia*
[*]*sbelim@mail.ru*

Abstract. All steganographic methods are focused on a specific container file format. Text documents with markup are the most difficult object for steganography methods. The article suggests a model for embedding structured text documents in control tags. The model uses the document tree structure and embeds into free leaf nodes. This approach adds hidden data that does not affect the display of the document. Two steganographic methods are implemented based on this model. The first method embeds hidden data into html document tags. The embedding method adds underplayed tags and style classes to free leaf nodes. The hidden data extraction method uses the embedding identifier. This role is played by the name of the new class. The name generation algorithm is based on the embedding key and hash function. The format of the identifiers matches the format of the source document names. This naming method allows the hidden message blocks to be randomly allocated to free leaf nodes. The second method embeds steganographic inserts into xml documents. Hidden data is added to the free leaf node attributes. The method requires two new attributes to execute. The optional structure describes both attributes. The format of this structure is indistinguishable from the structures present in the document. The embedding identifier is also based on the embedding key and the embedded block number. The data view uses an encryption algorithm with an additional key. Both methods use embedded data masking to counteract source code steganalysis. Steganalysis of such methods has exponential algorithmic complexity, so both methods are only applicable to large files.

Keywords: steganography, embedding model, hierarchical file, xml format, html format

For citation: Belim S.V., Munko S., Belim S.Yu. Steganographic embedding model in files with hierarchical structure. *Prikladnaya informatika*=Journal of Applied Informatics, 2025, vol.20, no.1, pp.125-139 (in Russian). DOI: 10.37791/2687-0649-2025-20-1-125-139

Введение

Стеганографические методы встраивания данных в легальные файлы приобретают всё большее распространение как дополнительные механизмы защиты информации. Основной целью этих методов ставится сокрытие самого факта наличия сообщения. Стеганографическая вставка формируется в файле-контейнере в результате изменения исходных данных. При этом выдвигается требование необнаружимости встроенных данных при штатном использовании контейнера. Второе требование сводится к устойчивости к стегоанализу. При стегоанализе исследуется исходных текст файла с целью обнаружения аномалий структуры и значения данных, свидетельствующих об искусственном изменении его содержания. В связи с этим стеганографические методы всегда жестко связаны со структурой файла, который используется в качестве контейнера.

Наибольшее распространение получили стеганографические алгоритмы встраивания данных в мультимедийные файлы [1–4]. Эти файлы включают в себя большое количество информации,