factor [h]

Функциональные характеристики

Factor H 1.0

Наименование программы	Factor H
Версия документа	1.0
Дата	18.10.2026

1. Назначение *3*
2. Состав модулей *3*
2.1. Модуль создания копий **3**
2.2. Модуль распознавания маркировки **3**
2.3. Роли пользователей **4**
3. Маркируемые материалы *4*
3.1. Поддерживаемые форматы **4**
4. Связь маркировки с получателем *5*
4.1. Ключ маркировки **5**
4.2. Подготовка копий **5**
5. Проведение расследования по образцу *5*
6. Формируемые результаты *5*
6.1. Результат идентификации **5**
6.2. Доказательная база **6**

1. Назначение

Factor H - это система невидимой маркировки документов и трассировки источника утечки. Для каждого получателя из реестра система выпускает индивидуальную копию исходного документа с встроенной невидимой маркировкой. Если копия или её фрагмент (файл, скриншот, фотография, скопированный текст) оказывается за пределами контролируемого периметра, система по совокупности маркеров устанавливает, кому именно была выдана эта копия.

Ключевые свойства:

- маркировка невидима и не меняет смысл и содержимое документа;
- опознание работает не только по исходному файлу, но и по скриншоту, фотографии экрана и даже по скопированному тексту;
- система самодостаточна: для опознания нужны только исследуемый образец и реестр получателей; теневые копии оригиналов не хранятся;
- при недостатке сигнала система честно отказывается от вывода, а не называет случайного получателя (отсутствие ложных обвинений).

2. Состав модулей

Система поставляется как единый веб-сервис и функционально состоит из двух прикладных модулей, доступных оператору через веб-интерфейс: модуль создания копий и модуль распознавания маркировки.

2.1. Модуль создания копий

Выпускает индивидуально маркированные копии исходного документа, по одной на каждого получателя из реестра.

- принимает на вход исходный документ (PDF, DOCX, XLSX, PPTX) и список получателей;
- встраивает в каждую копию невидимую маркировку, однозначно связанную с конкретным получателем, одновременно в несколько независимых каналов;
- не изменяет видимый текст, состав и смысл документа;
- упаковывает готовые копии в архив для выдачи получателям.

2.2. Модуль распознавания маркировки

Устанавливает получателя(источник) по предъявленному образцу (расследование утечки).

- принимает на вход образец: файл копии, скриншот, фотографию экрана или скопированный текст;
- считывает маркировку послойно, по убыванию надёжности;

- сопоставляет измеренную метку с кодами получателей из реестра текущего оператора;
- выдаёт вердикт с указанием получателя и достоверностью, либо честный отказ «источник не определён» при недостатке сигнала;
- формирует доказательную базу (карта совпадения знаков, рейтинг кандидатов, числовые показатели).

Оба модуля опираются на общий реестр получателей и общее ядро маркировки и опознания. Доступ разделён ролями: оператор выпускает копии и ведёт расследования в пределах своего изолированного реестра, администратор управляет системными пользователями.

2.3. Роли пользователей

Система разграничивает доступ по ролям:

Роль	Доступные функции
Оператор	Работа в пределах своего изолированного реестра получателей: ведение реестра получателей, выпуск маркированных копий, проведение расследований, просмотр результатов и доказательной базы
Администратор	Управление системными пользователями: создание и ведение учётных записей операторов и администраторов, назначение ролей, смена паролей.

Реестры получателей разных операторов изолированы: расследование выполняется только в пределах реестра текущего оператора.

3. Маркируемые материалы

3.1. Поддерживаемые форматы

- Выпуск копий (маркировка): PDF, DOCX, XLSX, PPTX.
- Распознавание маркировки (на входе): те же форматы плюс изображения (PNG, JPG/JPEG, TIFF/TIF, BMP) и вставленный текст.

3.2. Независимые слои защиты

В каждой копии сосуществуют несколько независимых каналов, в которые добавлена маркировка. Если один утрачен (печать, пересохранение, копирование текста), опознание идёт по оставшимся.

4. Связь маркировки с получателем

4.1. Ключ маркировки

Несущая модулируется индивидуальным кодом получателя. Код детерминировано вычисляется из атрибута получателя (email, user id, user account name и т. п.), поэтому у каждого получателя свой уникальный набор маркеров. Атрибут получателя является единственным ключом для маркировки и хранится в реестре получателей оператора вместе с техническим идентификатором и именем.

4.2. Подготовка копий

Оператор загружает исходный документ, и система выпускает по одной копии на каждого получателя из своего реестра. Каждая копия несёт код соответствующего получателя во всех слоях. Копии упаковываются в архив для выдачи.

5. Проведение расследования по образцу

На вход подаётся образец (файл копии, скриншот, фотография или вставленный текст). Область поиска ограничивается реестром получателей текущего оператора (реестры получателей изолированы). Опознание выполняется послойно, по убыванию надёжности слоев защиты.

Принцип опознания. Для каждого кандидата выполняется корреляция измеренного сигнала метки с его кодом и вычисляется z-оценка, характеризующая степень превосходства наилучшего кандидата над фоном случайных кодов. Порог принятия решения адаптируется к числу получателей (при увеличении реестра порог повышается) таким образом, чтобы вероятность ложной идентификации сохранялась на постоянном низком уровне независимо от размера реестра.

6. Формируемые результаты

6.1. Результат идентификации

- «Источник установлен» — указывается конкретный получатель (ФИО, email, технический идентификатор), достоверность (в процентах), дата выдачи и перечень сработавших факторов.
- «Источник не определён» — сигнала недостаточно для достоверного вывода (слишком короткий или некачественный образец, документ без текстового слоя и т.п.). В этом случае получатель не указывается; система не формирует предположительный вывод при отсутствии достаточного сигнала.

6.2. Доказательная база

К результату идентификации прилагаются следующие материалы, формируемые системой и отображаемые в интерфейсе расследования:

- **Соответствие по получателям** — соответствие образца получателям реестра. У истинного получателя соответствие многократно выше, чем у остальных; список ограничен первыми 20 кандидатами с наибольшим рейтингом.
- **Детализация совпадения** — по каждому измеренному сигналу (несущей) показано, совпал ли его знак с кодом названного получателя: отображаются доли «совпало» и «расхождение», а также итоговый процент совпадения.
- **Объём образца** — число несущих, по которым проводилось измерение.
- **Сигнал / порог** — измеренный уровень сигнала и адаптивный порог принятия решения; соотношение показывает, насколько сигнал превышает порог.
- **Совпадение** — итоговая доля совпавших знаков в процентах.
- **Рейтинг получателей** — упорядоченный по уровню список кандидатов с указанием ID получателя, ФИО и должности, email и числовой оценки; у истинного получателя оценка существенно выше, чем у остальных.
- **О заключении** — пояснение, что заключение опирается на машинно-проверяемое доказательство: соответствие атрибутов исследуемого образца атрибутам всего пула получателей. Порог адаптируется к числу получателей, а достоверность рассчитывается как вероятность ошибки при данном уровне сигнала.