



Раер Михаил Геннадьевич
Фишман Григорий Михайлович

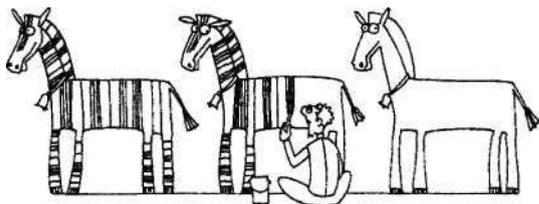
ПОЛОСАТЫЙ ПРОПУСК

Данная публикация является результатом наших исследований. Авторы не использовали никакие источники закрытой информации. При желании любой из вас мог проделать ту же работу и прийти к аналогичным результатам. Авторы не несут никакой ответственности за использование данного материала кем-либо в противозаконных целях и за материальные убытки, понесенные в результате этого использования. Это исследование можно использовать в любых законных целях без ограничений, но с обязательной ссылкой на авторов и на наш сайт (<http://tickers.da.gu>). Если у вас есть что добавить или уточнить, также ждем вас на вышеуказанном сайте. Наш e-mail: tickers@yandex.ru

дорожник инженер Давид Коллинз. В 50-х годах он окончил Массачусетский технологический институт и пошел работать на Пенсильванскую железную дорогу. Там он занимался сортировкой вагонов. Это был тяжкий труд. Вагоны нужно было пересчитать, проконтролировать по документам каждый вагон, определить путь следования. Тогда он придумал записывать номера вагонов кодом, состоящим из красных и синих полос. Длина такого кода достигала полуметра...

...Шло время. Полосы превратились в черно-белые, штрих-код умещается на билете шириной 4,5 см, да и цели перестали быть такими бескорыстными и благородными...

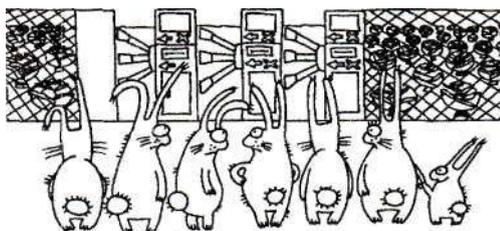
ИСТОРИЧЕСКАЯ СПРАВКА



Сейчас штрих-коды широко используются для индексации различных товаров. Их можно встретить на обертках продуктов питания, на книгах, тетрадях, библиотечных читательских билетах и т.д. Мы же расскажем об их использовании на железнодорожных билетах.

Ирония судьбы состоит в том, что штриховое кодирование изобрел железно-

МПС И АВСТРАЛИЯ ПРОТИВ КРОЛИКОВ



Совсем недавно МПС жилось плохо. Ей мешали собирать капусту многие зайчики и кролики. Поэтому, для сохранения бизнеса, МПС поставила перед собой задачу: оградить свое поле высокой стеной и сделать систему платных пропусков для обитателей леса. Эта система должна была удовлетворять следующим требованиям:

1. Позволять проходить как по абонеентам, так и по одноразовым билетам.
2. Автоматически контролировать доступ пассажиров на платформы.
3. Обеспечить низкую себестоимость билета (стоимость клочка бумаги).
4. Билет должен быть защищен от подделок.
5. По билету можно пройти, только если он датирован текущим днем.
6. По билету любой человек должен суметь узнать дату билета, исходную и конечную зоны маршрута, направление, цену и тип билета. Эта информация должна быть доступна контролерам и пассажирам без использования технических средств.
7. На билете должна быть информация о номере кассы и номере билета для проверки фальсификаций, повторного использования и контроля денежных средств.

На возможностях этой системы, видимо, было решено сэкономить: была сформулирована концепция автономности станций, по которой, например, по билету, купленному и использованному на одном вокзале, можно пройти на другом вокзале той же зоны.

Теперь рассмотрим разные варианты реализации одноразовых билетов:

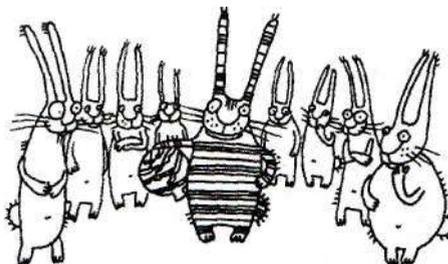
1. Магнитная или смарт-карта отпадают по п. 3 и п. 6.
2. Жетоны отпадают по п. 5 и п. 6.
3. Бумажный билет с водяными знаками и напечатанными данными отпадает по п. 2,



Рисунок 1.

Единственное, что осталось - это клочок бумаги, с напечатанными на нем данными и *секретным* штрих-кодом, по которому турникет сможет определять параметры билета. Именно такой вариант и был выбран разработчиками системы (рисунок 1).

С ЧЕМ ЕДЯТ ШТРИХ-КОД



Штрих-код - это не магнитная или смарт-карта, это всего лишь последовательность вертикальных полосок, которую может напечатать каждый. Поэтому непонятно, почему билет так слабо защищен!

Существует несколько разных форматов штрих-кодов. Они бывают линейные и двумерные. Нас будет интересовать линейный код *ITF (Interleaved 2 из 5)*. Код состоит из последовательности чередующихся черных и белых вертикальных полосок, начинающейся и заканчивающейся черной. Полоски бывают широкие (будем обозначать 1) и узкие (будем обозначать 0). *ITF* предполагает наличие стартового и стопового (финишного) символов, которые кодируются 0000 (то есть четыре узкие полоски) и 100 (одна широкая и две узкие полоски), соответственно. Между этими символами и расположена полезная информация.

Свое название код получил, исходя из того, что информация в нем кодируется расположением двух широких полосок среди пяти. Применяя элементарную комбинаторику (число сочетаний), получаем, что таким блоком можно закодировать $C_2^5 = 10$ различных значений, то есть все цифры от 0 до 9 (попробуйте сами, в качестве упражнения, расписать все десять возможных вариантов расстановки двух единиц среди трех

нулей). Interleaved (перемежающий) же он потому, что рассматриваются отдельно черные и отдельно белые полоски. Первой записывается цифра, образованная черными полосками (рисунок 2), Алфавит кода можно найти на <http://www.sbarcode.com/encoding/int25.shtml> или на нашем сайте <http://tickers.da.ro>.

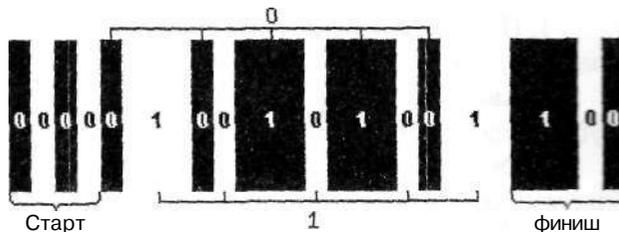
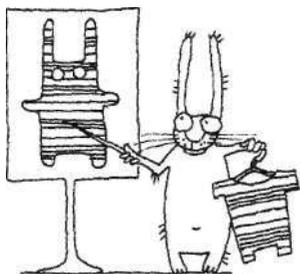


Рисунок 2. Пример чтения штрих-кода. Закодирована последовательность **01**.

КРОЛИКИ РЕШИЛИ ПОДУМАТЬ...



Штрих-код железнодорожного билета состоит из восьмидесяти семи полосок. Семь из них это стартовый и столовый (финишный) символы.

Остается восемьдесят, которые несут информацию. Делим на пять и получаем, что там записано шестнадцать десятичных цифр.

Что же это за шестнадцать цифр? Что там записано? Может быть, это номер записи в базе данных проданных билетов? Нет, этот вариант отпадает из-за автономности касс. Там закодирована вся важная информация о билете (см. таблицу 1)!

Сразу возникают следующие вопросы:
Вопрос: Почему не кодируются все цифры номеров касс и билета?

Ответ: Ответ прост - А зачем? Этого и так вполне достаточно. Тем более, ощущается явная нехватка места для кодирования на билете более чем шестнадцати цифр.

Вопрос: Какая позиция в штрих-коде у каждого поля?

Ответ: Пока рано об этом говорить. Во-первых, этого недостаточно знать, по той причине, что турникет пропускает билеты только сегодняшней даты, а поля даты у нас в таблице не отражены.

Во-вторых, расположение полей в штрих-коде (*перестановка*) меняется каждый день.

И, в-третьих, к каждому значению поля каждый день прибавляется по модулю десять (берется остаток от деления суммы на десять) определенное для данного дня и

Таблица 1. Цифры, закодированные в штрих-коде

Обозначение десятичной цифры (<i>поля</i>) Нумерация производится справа налево	Значение поля	Пример с рисунка 1
К0, К1, К2	Последние три цифры номера кассы (строка «БПМФ...» в билете).	К0 = 4; К1 = 5; К2 = 7
Н0, Н1, Н2, Н3	Последние четыре цифры номера билета (строка «БИЛЕТ N...» в билете).	Н0 = 0; Н1 = 7; Н2 = 1; Н3 = 4
Z0, Z1	Зона прибытия	Z0 = 0; Z1 = 0
Z2, Z3	Зона отправления	Z2 = 2; Z3 = 0
D	Тип билета и направление (расшифровка далее)	
CZ0, CN0, CK0, CN1	«Четыре константы» (о них будет сказано позже)

Таблица 2.

K1	Z2	N2	D	CZO	ZO	CM	N1	K2	N3	Z1	Z3	CNO	NO	CKO	KO
8	1	5	4	1	2	0	5	0	6	3	2	0	2	4	0

данного поля число (которое мы называем *смещением*).

Рассмотрим *перестановку* и *смещения* определенной даты. Сделаем это, например, для 17 марта 2001 г. (см. таблицу 2).

Всю таблицу мы называем *маской*.

Первую строчку - *перестановкой*.

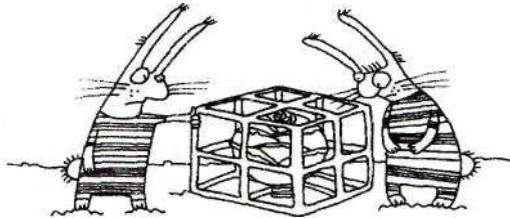
Вторую - *вектором смещения*.

Давайте подведем первый итог:

Чтобы узнать шифр, то есть научиться по штрих-коду узнавать все данные о билете, и наоборот, генерировать штрих-код по любым желаемым данным, достаточно знать соответствующую дату маски.

Таким образом, по приведенной маске можно создавать и распознавать абсолютно любой билет, датированный 17-м марта 2001 года...

ОНО БЫЛО ИХ СЕКРЕТОМ,..



ПРЕЛЮДИЯ

Все началось в марте 2001-го... На большинстве лекций было скучно, дома печально, а в аквариуме (так называется наш компьютерный класс) грустно. И вдруг, как гром среди ясного неба, Миша предложил новый увлекательный проект...

И мы почти всей группой стали ходить на вокзал, собирать билетки и анализировать их. Наш аквариум был сразу же перепрофилирован из места, в котором происходят великие битвы в Starcraft и сражения в шахматы по интернету с одноклассниками из соседней части аквы, в центр обработки билетов. Треть людей вносила их в базу данных, сфокусировав свое зрение на

гуще полосок, треть писала различные программы для их анализа, а оставшаяся треть сидела на лекциях, вглядываясь в бесконечные потоки цифр. Постепенно начала выявляться некая логическая структура. И вдруг на одной из лекций наш чемпион Александр Штучкин уловил определенную закономерность (вспомните фильм «Игры разума» со сценой в Пентагоне). Это был поворотный момент, после которого наши исследования перешли на новый качественный уровень. Однако вскоре наступила сессия, и приоритеты сдвинулись. Летом 2001-го была написана программа, которая по известным нескольким билетам одной даты выдавала все варианты масок, отбрасывая те, которые не подчиняются *парному закону*, о котором вы узнаете далее. Вскоре был сделан новый (и на данный момент последний) качественный скачок, и в программу была добавлена функция, благодаря которой стало достаточно вводить лишь один билет для каждой даты.

ЧЕТЫРЕ КОНСТАНТЫ

Константы так названы потому, что их значение не изменяется в течение дня. Возникает даже здоровое подозрение о том, что это дата... Однако тщательные исследования показали, что одна из констант все-таки меняется. Но мы заметили это слишком поздно, и в силу этого поле сохранило свое название. В науке, а особенно в физике, существует масса примеров, когда первоначальное неправильное представление о чем-либо приводило к неправильной терминологии, которая продолжает использоваться в силу привычки. Мы же, подобно физикам-экспериментаторам, следуя их дурному примеру (а он, как всем известно, заразителен), сохранили за этим полем название *константа*.

Перечислим константы: CZO, CN0, CKO и та самая заблудившаяся потеряшка,

псевдоконстанта CN1, Как корабль назовешь, так он и поплывет. Внимательный читатель обратил внимание на то, что в названия констант входят имена других полей. Для чего это сделано, будет объяснено дальше, но уже в следующем абзаце...

ТАНЕЦ ДАТСКОГО КОРОЛЯ ИЛИ «ЗАКОН ПАР»

Изучая первые полученные маски, мы обнаружили, что расположение полей день изо дня «скачет». И вскоре, изучая эти «скачки», заметили, что все перестановки осуществляются только парами. То есть существуют восемь неразлучных пар полей, которые и переставляются. Чем руководствовались разработчики, нам непонятно. Вряд ли их целью было нас запутать, потому что так нам стало гораздо легче - мы смогли различать константы. Просто назвали константу именем поля, которое с ней в одной связке.

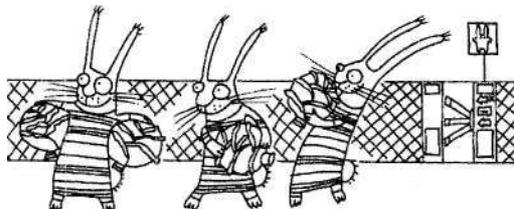
Далее мы призадумались над тем, как осуществляются перестановки пар. Искали период, но безуспешно. Казалось бы, вот расположение пар повторяется на девятый день. Ждем еще «9 дней, 9 ночей», ожидая ту же маску... Как бы не так - другая. И только много месяцев спустя мы нашли период. Он оказался равным шестнадцати дням. Притом, что масок всего десять. То есть внутри периода тоже бывают совпадения масок. Это и вводило нас в заблуждение. Разработчики же добились своего - выиграли некоторое время.

«О ПОЛЕ D ЗАМОЛВИТЕ СЛОВО...»

В этом абзаце мы осветим существующие типы билетов и направления. Четыре типа билета имеют два направления, и еще два типа не имеют направлений. Первые четыре - это *ПОЛНЫЙ*, *ДЕТСКИЙ*, *ЛЬГОТНЫЙ* и *БЕСПЛАТНЫЙ*. Каждый из них может быть как *ТУДА* (->), так и *ТУДА И ОБРАТНО* (<->).

Остались типы, которые идентифицируют себя просто цифрой 9 и словом *абонемент*.

КОЧЕРЫЖКА ЦВЕТНОЙ КАПУСТЫ



Что же такое константы? Так как три константы в течение дня не изменяются, мы предположили, что это дата или функция от даты. Логично? Смысл четвертой константы остается для нас загадкой. На билетах она принимает за день всего два значения - «i» и «i + 1». Поэтому-то мы и зачислили по ошибке ее в ряды констант. Но на самом деле она может быть равна чему угодно, и валидность билета от нее не зависит.

А теперь, поговорим о том, что нам с Вами предстоит еще открыть. Поля не хранятся в штрих-коде в явном виде. К каждому полю в маске прибавляется по модулю десять определенное число, которое мы назвали смещением. Именно сумма реального значения поля и смещения записывается в штрих-код. Слава Богу, что смещения не меняются в течение дня. Таким образом, шестнадцать смещений в маске можно представить как шестнадцать функций от даты. Осталось их найти. Может, они периодичны? А может, просто удастся их экстраполировать? В любом случае, имея билет на какое-то число, зная его реальные характеристики, маску и значения полей в штрих-коде, можно без труда, с помощью операции вычитания узнать все шестнадцать смещений на сегодня. А потом, зная эти шестнадцать чисел, изготовить билет с любыми характеристиками, но опять же только на сегодня.

АВТОНОМНЫЙ ТУРНИКЕТ

Как турникет распознает дату билета? Может быть, так:

Билет попадает в турникет. Турникет предполагает, что билет сегодняшний. Считывает цифры в том месте, где сегодня дол-

жны лежать три константы (дата), вычитает смещение. И если полученное значение не совпало с сегодняшней датой, значит, его предположение неверное, и билет не сегодняшний. Далее он предполагает, что он вчерашний и т. д., пока не наткнется на совпадение предполагаемой и кодируемой в штрих-коде дат.

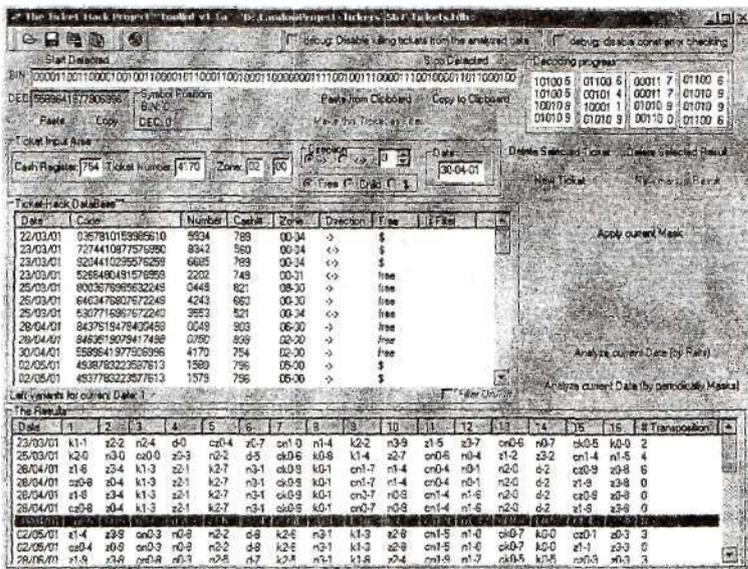
Загадкой всей системы осталась функция смещения от даты. Но мы думаем, что это всего лишь вопрос времени. Так что, держайте, «лед тронулся, господа присяжные заседатели...»

Особую благодарность выражаем Анатолию Абрамовичу Шалыто, который вдох-

новил нас на написание статей своими зажигательными лекциями о «SWITCH-технологии» (<http://is.ifmo.ru>).

Ссылки.

1. <http://tickers.da.ru> (Сайт, посвященный данной тематике)
2. tickers@yandex.ru (Авторы)
3. <http://www.sbarcode.com> (Сайт, посвященный штриховому кодированию, На этом же сайте можно найти демо-версию программы LabelBar для печати штрих-кодов)
4. <http://www.barcode.kiev.ua/history.html> (история штрихового кодирования)



Раер Михаил Геннадьевич, студент 3 курса факультета информационных технологий и программирования СПбГИТМО (ТУ).

Фишман Григорий Михайлович, студент 3 курса факультета информационных технологий и программирования СПбГИТМО (ТУ).

