SAND2014-15758D

Exceptional service in the national interest

Sandia National Laboratories

iOS Clone Detection Analysis

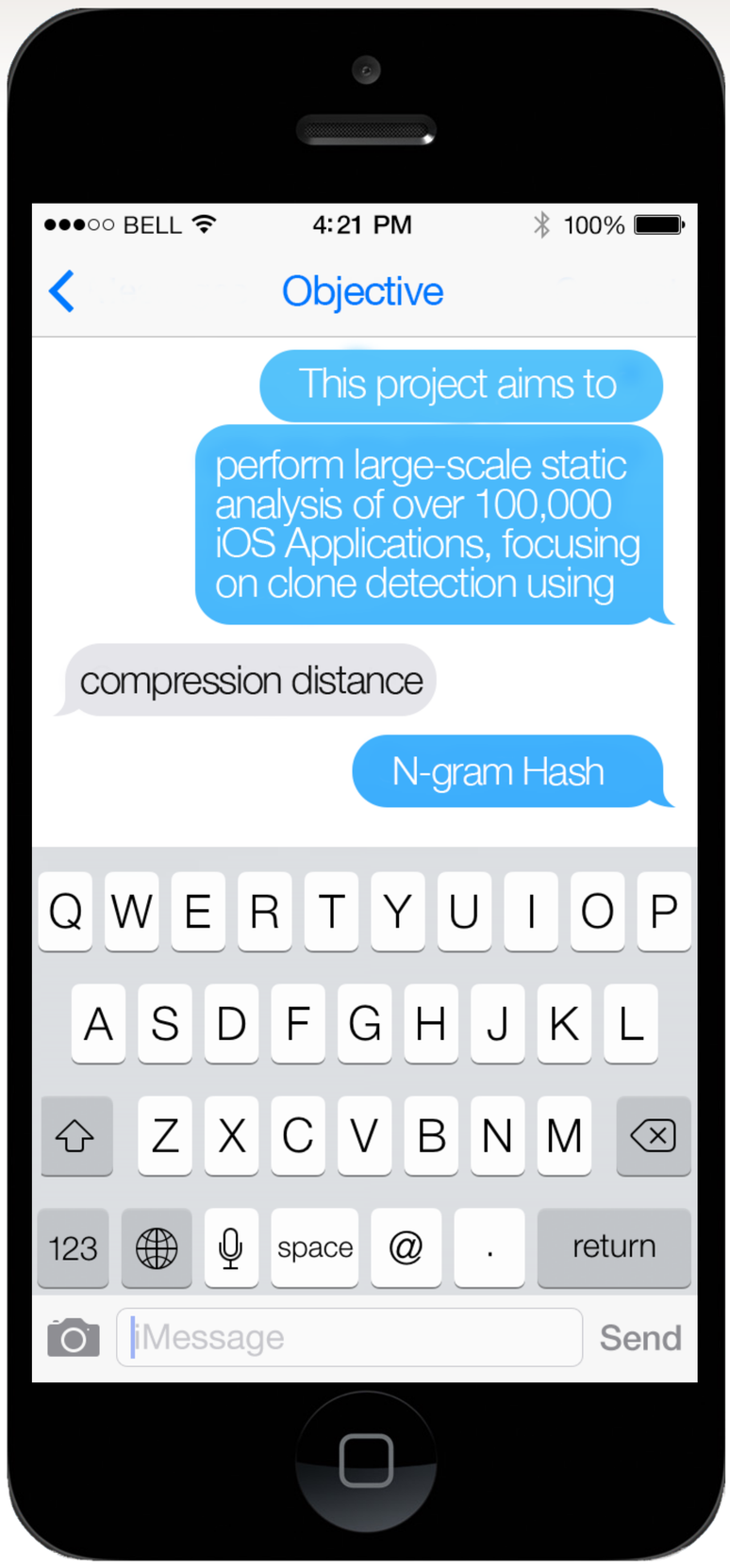
- Clone Detection
- Malware Analysis

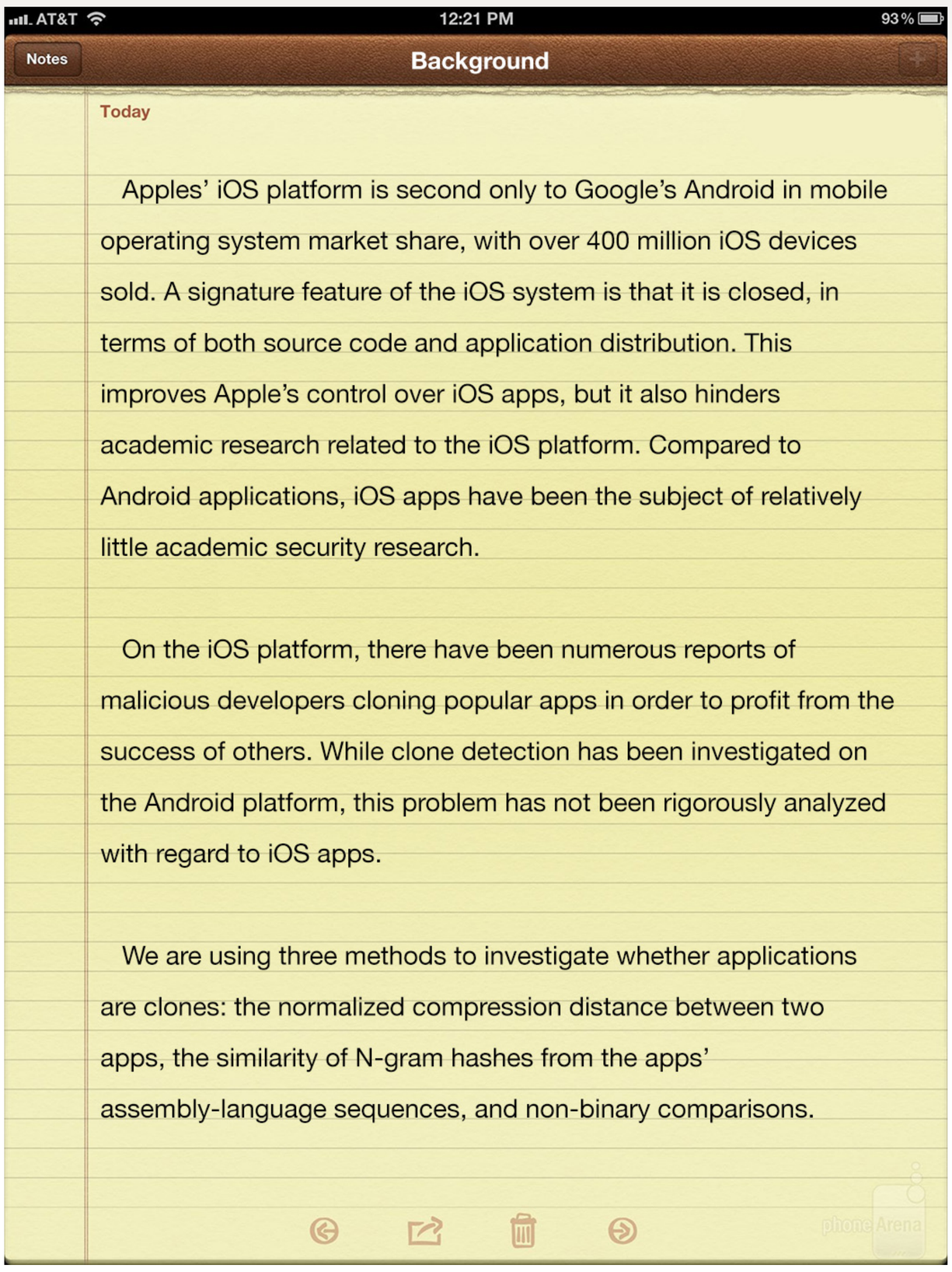
Michael Bierma, UC Davis, M.S. mhbierma@ucdavis.edu

Nicholas Ward, UC Berkeley, B.A. kingsyphax@berkeley.edu

Kevin Wu, UIUC, B.S. kcwu2@illinois.edu

Yung Ryn Choe, Sandia National Laboratories, yrchoe@sandia,gov





Compression Distance

If it's easier to compress the binaries of two apps together than separately--that is, if the compression of the concatenation of the apps is much smaller than the concatenation of the compressions of the apps--the apps are similar, suggesting cloning.

Method #1

Similarity of N-gram Hashes

The binaries' assembly-level instructions are split up into N-grams (groups of N consecutive instructions, where N is an integer). Fuzzy hashes are computed of each list of N-grams. Similarity of these fuzzy hashes indicates the apps have similar N-grams, so their code is similar.

Method #2

Non-binary Comparison

We compared images and other files to one another to check for cloning. We are planning to use a variety of methods such as decision trees, perceptual hash, and keypoint matching to compare images.

Method #3

ENERGY

National Nuclear Security Administration

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.
SAND No. 2011-XXXXP

Exceptional service in the national interest

Sandia National Laboratories

iOS Clone Detection Analysis

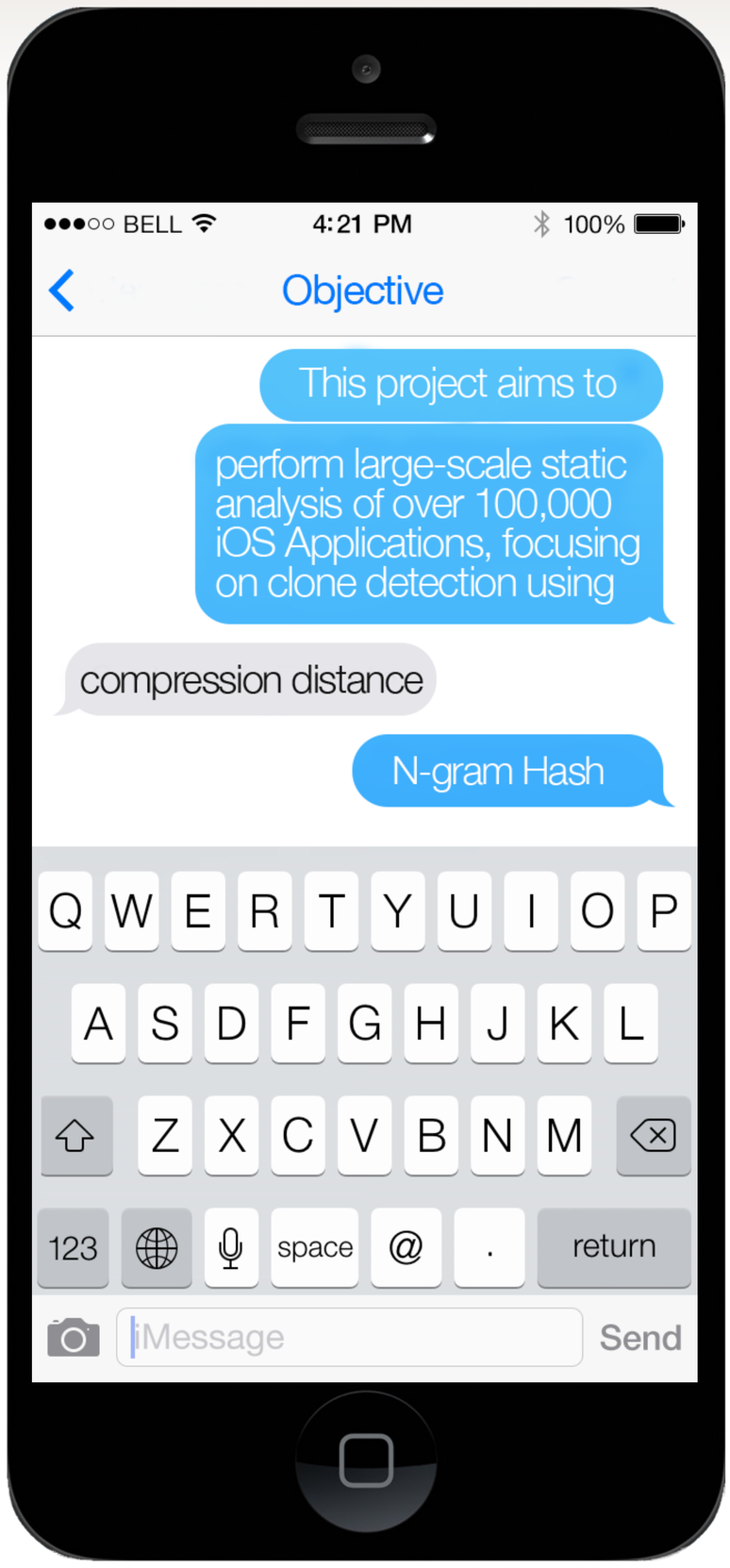
- Clone Detection
- Malware Analysis

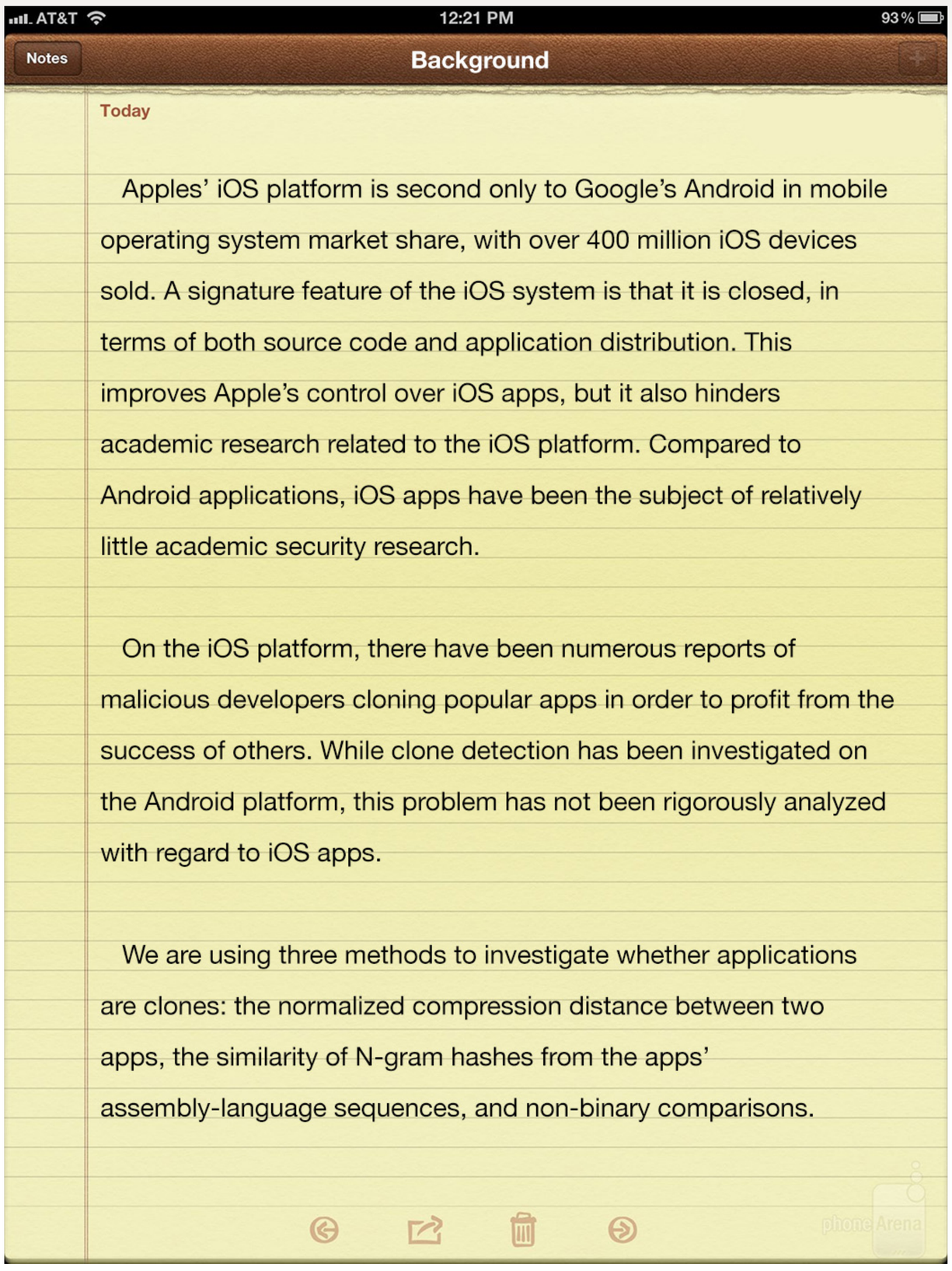
Michael Bierma, UC Davis, M.S. Comptuer Science, 2014

Nicholas Ward, UC Berkeley, B.S. Mathematics/Computer Science, 2018

Kevin Wu, UIUC, B.S. Computer Engineering, 2018

Mentor: Yung Ryn (Elisha) Choe, 8965, Information Assurance





Compression Distance

If it's easier to compress the binaries of two apps together than separately--that is, if the compression of the concatenation of the apps is much smaller than the concatenation of the compressions of the apps--the apps are similar, suggesting cloning.

Method #1

Similarity of N-gram Hashes

The binaries' assembly-level instructions are split up into N-grams (groups of N consecutive instructions, where N is an integer). Fuzzy hashes are computed of each list of N-grams. Similarity of these fuzzy hashes indicates the apps have similar N-grams, so their code is similar.

Method #2

Non-binary Comparison

We compared images and other files to one another to check for cloning. We are planning to use a variety of methods such as decision trees, perceptual hash, and keypoint matching to compare images.

Method #3

ENERGY

National Nuclear Security Administration

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.
SAND No. 2011-XXXXP