All Your Base are Belong to US

Richard Ford
Harris Institute for Assured Information
Florida Institute of Technology
150 W. University Blvd.
Melbourne, FL 32901
rford@fit.edu

Liam M. Mayron
Harris Institute for Assured Information
Florida Institute of Technology
150 W. University Blvd.
Melbourne, FL 32901
lmayron@fit.edu

ABSTRACT

In this paper we examine an important example of where a decision designed to improve security has quite the opposite effect due to the longevity of the decision's side effects. The primary example we use to illustrate our point is the deliberate obfuscation and alteration of satellite imagery available online, in an attempt to deprive attackers of sensitive information about certain installations. We will demonstrate that such image modification techniques are short sighted and counterproductive: better security would have been provided if the concerned parties had simply done nothing. This result illustrates the importance of examining security decisions from a temporal perspective, accounting for inevitable progression of technology. Furthermore, we believe this paper illustrates an oft overlooked class of security problems that present non-reversible challenges to the defender long after the questionable security decision has been reversed.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous

General Terms

Image Manipulation, Censorship, Mapping

Keywords

Image Manipulation, Security through Obscurity, Geospatial References

1. INTRODUCTION

Satellite imagery has been available to the military for a relatively short time. The first recorded satellite imagery was taken from a US-operated V2 rocket in 1946 [15]. Since then, satellite imagery has progressed from government-only purposes to a staple of the web, with a variety of imagery becoming available.

When high resolution satellite imagery first became publicly accessible, both governments and companies expressed some concern about its potential impact on privacy and security – after all, such imagery would provide attackers with a potentially critical advantage when conducting espionage or planning an attack. In response to these concerns, several images on these servers have been modified to obscure or alter certain details. Modifications range from obvious (blurring or obscuration) to subtle (the pasting of other images in place of the actual imagery), in the name of "security".

In this paper, we take the position that this response was, at best, counterproductive. Using simple techniques, we demonstrate that image alteration serves no purpose, and in fact draws attention to locations that might otherwise remain fairly anonymous. While this may seem obvious in the context of satellite imagery, it highlights an important question that defenders should ask themselves when examining countermeasures.

This illustration is but an example of a larger class of problems. Typically, when we consider a security decision (such as deploying a firewall, or installing antivirus software) the decision is temporally limited; that is, if we determine that there was a "better way" post fact, we can reverse our decision with few negative consequences. We argue that in fact there is a large and growing class of decisions related to security that are not easily reversible in totality, and that have long term consequences for the defender even after being addressed. Thus, the premise is that choices must be made accounting not only for the needs of today but also for the likely trajectory of technology well into the future. What at first glance seemed like a good idea in 2005 may cause us fairly predictable problems in 2012.

The remainder of this paper is organized as follows: Section 2 presents contextual information. Section 3 discusses the steady progress of relevant technology. In Section 4 an overview of techniques for obscuring aerial imagery is presented while Section 5 presented techniques for detecting manipulation. A case study exploring the automated detection of obscured locations in aerial imagery appears in Section 6. A discussion of the implications of using an outdated security paradigm is in Section 7. Finally, concluding thoughts are offered in Section 8.

NSPW'12, September 18–21, 2012, Bertinoro, Italy.

2. WELL-KNOWN EXAMPLES

Perhaps the discipline with the best understanding of long-lifetime security problems is encryption. Typically, when data are encrypted, the defender is well-aware that the encryption scheme used is not unbreakable. Given sufficient time, it is well known that most of our common encryption algorithms (save one time pads) can be broken by brute force. Algorithm choice and key length often directly impact this time — longer keys have a larger key space, and therefore take longer to crack.

When choosing how to deal with encrypted data, defenders are aware of the fact that encryption is not bullet proof: it simply buys time and costs the attacker computational resources to break. Thus, the level of encryption used depends on the temporal needs of the defender. For example, if the best known current technology would take ten years to brute force a session key for a website that is designed to secure a credit card number, the level of encryption is strong enough for practical purposes. Conversely, when encrypting a document that is to be sent across a public channel, this decryption cost may not be enough if the consequences of leaking that document are still important in ten years time. That is, the value of the information at the time it is likely to be compromised can be used as a gauge to determine how to secure it. Some information has long term sensitivity, some not. The defensive measure used depends on this as, especially as it relates to encryption, time is the currency of security.

Even in this well trodden field, we believe that a potentially serious issue may be brewing with respect to encryption algorithms. As the development of a usable quantum computer inches forward, the lifetime of some of our more common cryptographic algorithms is potentially being shortened. Given that there are several usable algorithms believed to be quantum resistant (for an overview, see [13]), one has to question the sense of ignoring a seemingly inevitable "quantum leap" in technologies. The time required to switch our computing ecosystem from existing encryption algorithms is not small — in private conversation an estimate of one or two decades is not uncommon — and the time to build a usable key-breaking machine is unknown, but possibly of the same order. By continuing to roll out systems that use "old" algorithms, defenders are at least potentially ignoring a technological paradigm shift... that is, in a decade's time will we may realize that we witnessed, *en masse*, an example of the class of problems we are illustrating here.

While the encryption discussion is well known, society as a whole is still feeling its way around other information release issues. Perhaps the largest concern at this time is the issue of data aggregation across multiple sites. Search providers, such as Google hold terabytes of data related to individual search results; such data can be used to identify users who have not chosen to share their real world identity directly with Google (for a good example of this, see the controversy regarding the inadvertent release of AOL search data [1]). While search aggregators are not common yet – at least in terms of mindshare – the general public is starting to see the tip of the iceberg with the availability of applications like "Girls Around Me".

Similarly, many people post personal information freely on social networking sites such as Facebook and Google+. This information is often deeply revealing, and can be used in the short term to leverage important information that has immediate value (see [6] for an example). However, this information also has long term value. While we have yet to see a presidential election hinge on information posted by a candidate while at college, such an occurrence is inevitable in a world where information never dies. Such problems have led futurists to predict that people will be able to "wash" their identities later in life, changing their names and connections to lose distressing videos of times (mis)spent at college. One of the subtle changes that technology has brought is that the modern universe has a memory; sharing hard copies of some silly pictures of college days with friends thirty years ago is a very limited release. Doing so online is entirely another matter.

In the preceding discussion, we have given an example of problems that exist after their "fix" has been put in place. For example, for the avid Facebook sharer who changes their ways, stopping sharing does not "bring back" information carelessly posted; it remains, in theory, forever. Similarly, if the switch to quantum resistant algorithms is made too late, untold damage will be done to financial systems that are incapable of changing safeguarding technologies with sufficient speed. While we typically think of these long term problems as fairly narrow or obvious, we argue in the remainder of this paper that they exist more commonly than we think. In order to illustrate this, we now turn our attention to the practice of redacting or modifying satellite imagery in online mapping programs.

3. TEMPUS FUGIT

When Google first launched its mapping service in 2005 the world was introduced to the ready public availability of high-resolution satellite imagery. While the Google site was not the first site to ever offer such images, it quickly became well known, and the launch of the site caused some concern amongst "traditional" security groups, who were worried about the ability of attackers to use the imagery to plan attacks against a facility.

The quality of the imagery available is absolutely outstanding. The main imagery source for Google Maps is GeoEye-1 or IKONOS. According to GeoEye's website [7], GeoEye-1 has a panchromatic resolution of 0.41x0.41 meters, though the maximum image resolution provided to non-government sources is 20in (0.51m). A planned commercial launch in 2013 will launch GeoEye-2, which is reportedly capable of 0.25m imagery. The ready availability of such high-resolution images are potentially of significant concern to the US government... after all an attacker can easily use these for espionage or attack planning.

While the preceding sounds right, there are issues with the logic. The defense is based upon an assumption that either the attacker is looking for a specific object that can be identified via an overhead image, or the attacker is already aware of the importance of a location.

Even though satellite imagery may be beyond the reach of terrorist organizations (and even several countries) if an airborne image of a target is desired, it is not difficult to get. Most readers would be forgiven for thinking that the skies in America are tightly controlled, but in fact for the vast majority of airspace, private pilots can fly wherever they wish. The absence of satellite imagery can be trivially made up for by a quick flight in a Cessna... even over locations that are carefully blurred in Google Maps. If the attacker

(a) Google Maps Image

(b) VFR sectional chart

Figure 1: Corresponding images for the Oconoee Nuclear Station (note that the station is actually marked as a waypoint on the VFR chart making its presence obvious to pilots)

already knows the location, image modification is (as we will show) worse than useless.

For example, Oconoee Nuclear Station is blurred in Google maps – a screenshot of the current imagery is shown in Figure 1. A careful look at higher resolution clearly shows the blurring used around the plant. Ironically, the power plant is actually shown on VFR Sectional Charts (shown in Figure 1b) as a landmark. There is no restriction on overflight, and it would be trivial to simply ask a pilot from nearby Oconee County Regional Airport in Clemson (search for KCEU in any airport database) to take up a would-be attacker on a "joyride" with their camera. This situation – where obscured satellite imagery is not supported by a flight restricted area – occurs innumerable times in the US. We have to ask what the purpose of blurring the data actually was, in the minds of those who requested the image manipulation from Google; all that was accomplished was to draw attention to the site.

Multiple sources of aerial imagery are accessible, including Google Maps [8], Bing Maps [11], Yahoo! Maps [20], and OpenStreetMap [12], among others. Collaborative services such as Wikimapia [18] allow individuals to annotate regions of a map. Image hosting services including Flickr [5] allow querying by location. Twitter [16] can also search for tweets within a radius of a location. The GeoHack tool [19] provides a convenient method to access many location-related queries. In aggregate, this freely-available information can provide a wealth of knowledge about a location; all one needs are the GPS coordinates.

This leaves only the first case: that the defender believes that the unaltered image will allow the attacker to identify the area of interest. We argue the opposite: the image modification itself becomes the weak link that allows the attacker to find locations that are interesting.

4. MANIPULATION OF AERIAL SATELLITE IMAGERY

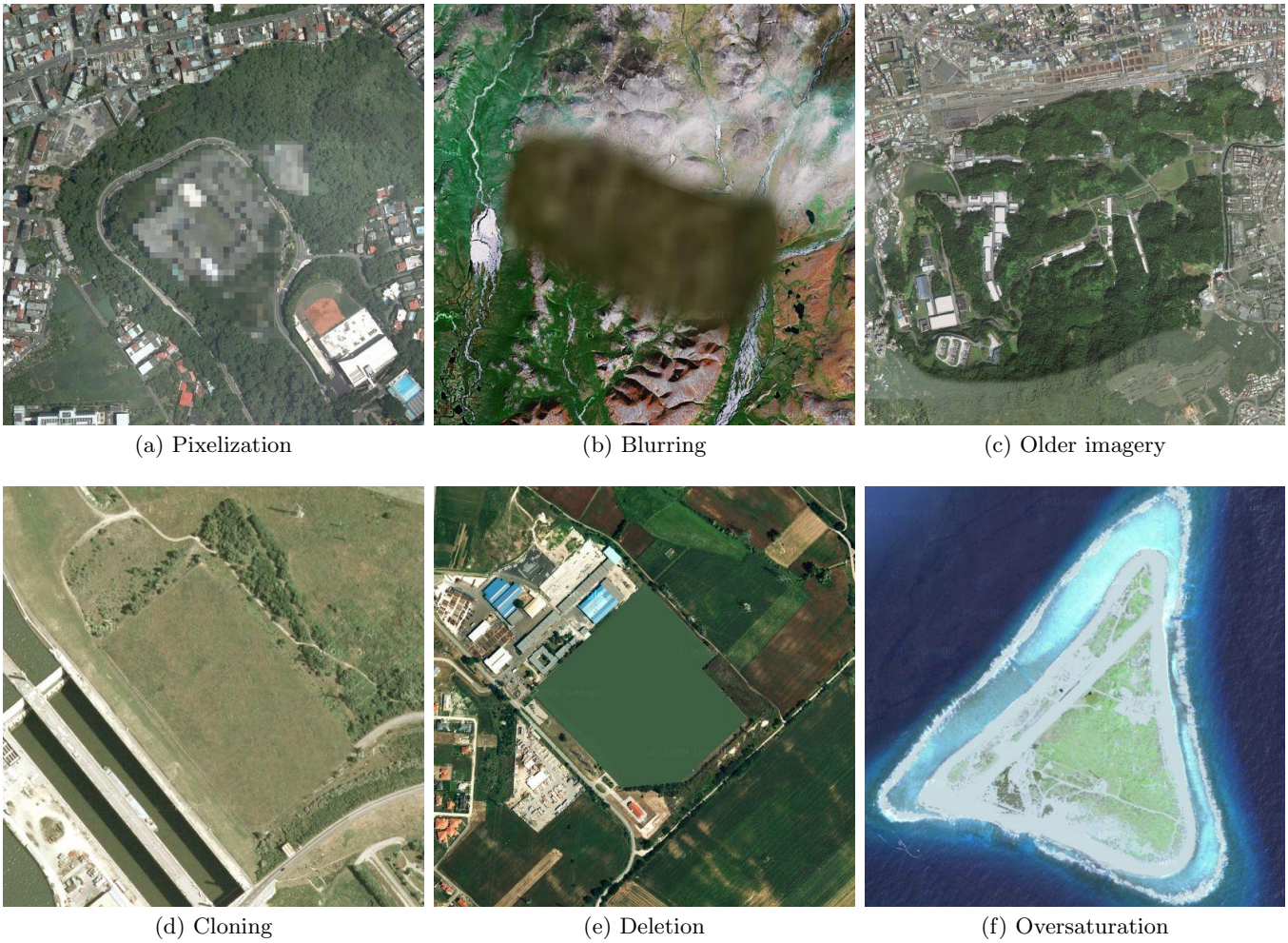
Known techniques employed to disguise aerial imagery are generally manipulations intended to obscure the nature, function, intricacies, or existence of an area. Manipulation techniques include:

- Pixelization (downsampling): pixelization is a technique that lowers the resolution of a particular area of an image (Figure 2a)
- Blurring: an area may have a low-pass filter applied to it, resulting in a blurred area (Figure 2b)
- Older imagery: antiquated imagery has the advantage that it can blend in more seamlessly with the surrounding environment, although it can still be possible to detect the manipulation due to differing resolutions and environmental conditions between the older and nearby newer imagery (Figure 2c)
- Cloning: map imagery from another location (or nearby) can be pasted on top of the region to be obscured. This can be particularly difficult to automatically detect, although physical properties such as lighting and shadows may be cues. (Figure 2d)
- Deletion: areas of an aerial image may be simply deleted (Figure 2e)
- Other manipulations: map areas may be manually-retouched or manipulated with a variety of other filters, such as posterization (reduction of a gradient into a single color) or overexposure (removing detail from bright areas of an image such as in Figure 2f)

Often, the primary objective of these manipulation techniques is to prevent ready interpretation of a location, not undetectability. However, as discussed in this paper, the detectability of an obscured area is an important component of its security.

5. DETECTING IMAGE MANIPULATION

The authenticity of an image may be evaluated by verifying a watermark or by a separate passive means [4]. Watermarks (similar to steganographic techniques) are difficult



(a) Pixelization (b) Blurring (c) Older imagery

(d) Cloning (e) Deletion (f) Oversaturation

Figure 2: Examples of censoring techniques in aerial map imagery

to remove artifacts that enable attribution of a image [14]. Watermarking is used to attribute map imagery to its content owner. However, it is not used to verify the integrity of the original image. Instead, a passive technique must be employed.

Passive techniques for detecting artificially-manipulated images may be [4]:

- Pixel-based: digital images are composed of pixels – the smallest unit of an image encapsulated by one or more numerical values corresponding to a grayscale or color component. Pixel-based techniques rely on image statistics to determine the presence of incongruent information.
- Format-based: lossy image compression formats, such as JPEG, introduce artifacts as part of their encoding schemes. For example, in JPEG, it is impossible to reconstruct the original image perfectly due to data loss during the quantization stage. If the manipulated image was based on a JPEG file and saved again as a JPEG, it may be possible to detect artifacts from repeating the quantization stage more than once.
- Camera-based: cameras have intrinsic qualities that affect the properties of the resulting image such as color and noise [4]. Disruptions to these properties may indicate tampering.
- Physically-based: lighting, shadows, and other elements from the world around us may not be consistent in manipulated imagery. For example, a house copied from another area may not match those surrounding it.
- Geometry-based: the position or relative size of an artificial object in an image may not be geometrically-plausible. Pixel-based techniques are the focus of the following case study, although other methods could be applied as well. For example, physically-based techniques may reveal the presence of cloned city block – reliable detection would complicate the application of cloning to censor map imagery (as it related to detectability).

6. CASE STUDY

A case study experiment was undertaken in order to explore the feasibility of automatically detecting (potentially unknown) obscured aerial imagery. The purpose of this test was not to be exhaustive, but to demonstrate the feasibility of automated detection using accessible tools. We limited

Table 1: Case study dataset characteristics

Region	Tiles	Terrain	Method
1	2250	Urban	Pixelization
2	2500	Mountain	Blurring
3	3600	Urban	Pixelization
4	10000	Lake	Blurring
5	10000	Urban	Blurring
6	7968	Forest	Pixelization

the experiment to two popular methods of obscuration: blurring and pixelization. Computational resources were modest (a single consumer laptop computer).

Although the earth is an oblate spheroid, it can be projected onto a flat surface. Online map imagery services generally allow browsing of aerial data by location and zoom level. Delivered data is divided into square tiles for convenient transmission and memory usage. A set of contiguous tiles will correspond to a region of the earth.

Patterns on the earth appear with regularity when captured in the visible spectrum. Image manipulation may introduce unnatural, but consistent artifacts. For example, older imagery may appear out of place or have different shadows. A popular method of obfuscating aerial map imagery is to blur of pixelate the sensitive area. This introduces a new texture that is consistent within the manipulated region, but not without it. As a result, the possibility for automatically detecting such manipulated regions merits exploration.

We did not seek to design a comprehensive method of automatically detecting manipulated imagery. Instead, our goal was to demonstrate a new paradigm by showing at least a single case where such manipulation could be detected. We present six examples where our proof-of-concept method is applied with success.

6.1 Methodology

A dataset consisting of six exemplar obscured locations was constructed. Data was collected automatically from a free online mapping service using a script to specify and download a range of map tiles (regions of the globe mapped to square images) at the highest zoom level available. The test regions consisted of between 2,250 and 10,000 tiles.

Imagery selected had a variety of characteristics (see Table 1). Terrain was either urban (containing many artificial structures, buildings, roads, etc.), mountain, lake (including a river), or forest.

The overall process of extracting regions that have been blurred or pixelated is illustrated by a block diagram (Figure 3). The process collects map tiles and computes a variety of statistical and texture-based metrics. Then, a rank (in terms of likelihood of being obscured) for each tile in an area is determine. An initial set of candidate locations are created. Morphological operations are used to consider the locations in context, remove isolated regions, and produce a better result. Finally, the mask is combined with the original image in order to determine the most likely obscured locations in an aerial terrain image.

A variety of texture-based metrics were computed for each tile in each image. Entropy is a measure of an image's complexity, derived directly from the gray-level pixel representation and defined by Equation 1, where r_i is the i^{th} possible gray level out L possible gray levels in an image and the function $p()$ determines the probability of that gray level's occurrence.

Table 2: Sorting scheme

Metric	Sorting
Entropy	Ascending
Contrast	Ascending
Correlation	Descending
Energy	Descending
Homogeneity	Descending

$$Entropy = -\sum i = 0^{L} - 1p(r_i)\log_2[p(r_i)] \quad (1)$$

The other metrics are derived from the *gray-level cooccurrence matrix* [10]. The i^{th}, j^{th} element of the gray-level cooccurrence matrix is denoted as $g(i,j)$ and corresponds to the number of occurrences of two gray levels with each other. The normalized version of this matrix is denoted by $N_g(i,j)$ and is derived by Equation 2 [10].

$$N_g(i,j) = \frac{g(i,j)}{\sum_i \sum_j g(i,j)} \quad (2)$$

The remaining metrics are contrast (Equation 3), correlation (Equation 4), energy (Equation 5), and homogeneity (Equation 6) [10]. For correlation, μ and θ are the mean and standard deviation of the sum of the denoted column or row, respectively.

$$Contrast = \sum_i \sum_j (i-j)^2 N_g(i,j) \quad (3)$$

$$Correlation = \frac{\sum_i \sum_j (i-\mu_i)(j-\mu_j)N_g(i,j)}{\theta_i \theta_j} \quad (4)$$

$$Energy = \sum_i \sum_j N_g^2(i,j) \quad (5)$$

$$Homogeneity = \sum_i \sum_j \frac{N_g(i,j)}{1+|i-j|} \quad (6)$$

The output of each metric is a scalar value that quantifies the particular texture characteristics for a given map tile. Given the texture metric for each tile in a map image, the tiles can then be sorted and ranked. Table 2 demonstrates the sorting scheme.

The various texture metrics produce different and not necessarily consistent results depending on the type of terrain they are applied to. A fusion of metrics was derived in order to provide stability and broader applicability. Score-level fusion was impractical due to the non-uniform range of the texture metrics. Instead, rank-level fusion was used. Rank-level fusion is appropriate as the goal is to derive a consensus rank for each tile. Thus, each tile in an image is assigned its rank within each sorted texture list. Then, the arithmetic mean is applied.

Let m and n respectively denote the rows and columns of tiles within a map area. Then, t can be defined as particular tile within the map area (with a value ranging from $0 \leqslant t < m \times n$. Let r be defined as a member of the set of possible texture metrics; $metric \in Metrics$ where $Metrics = entropy, contrast, correlation, energy, homogeneity$.

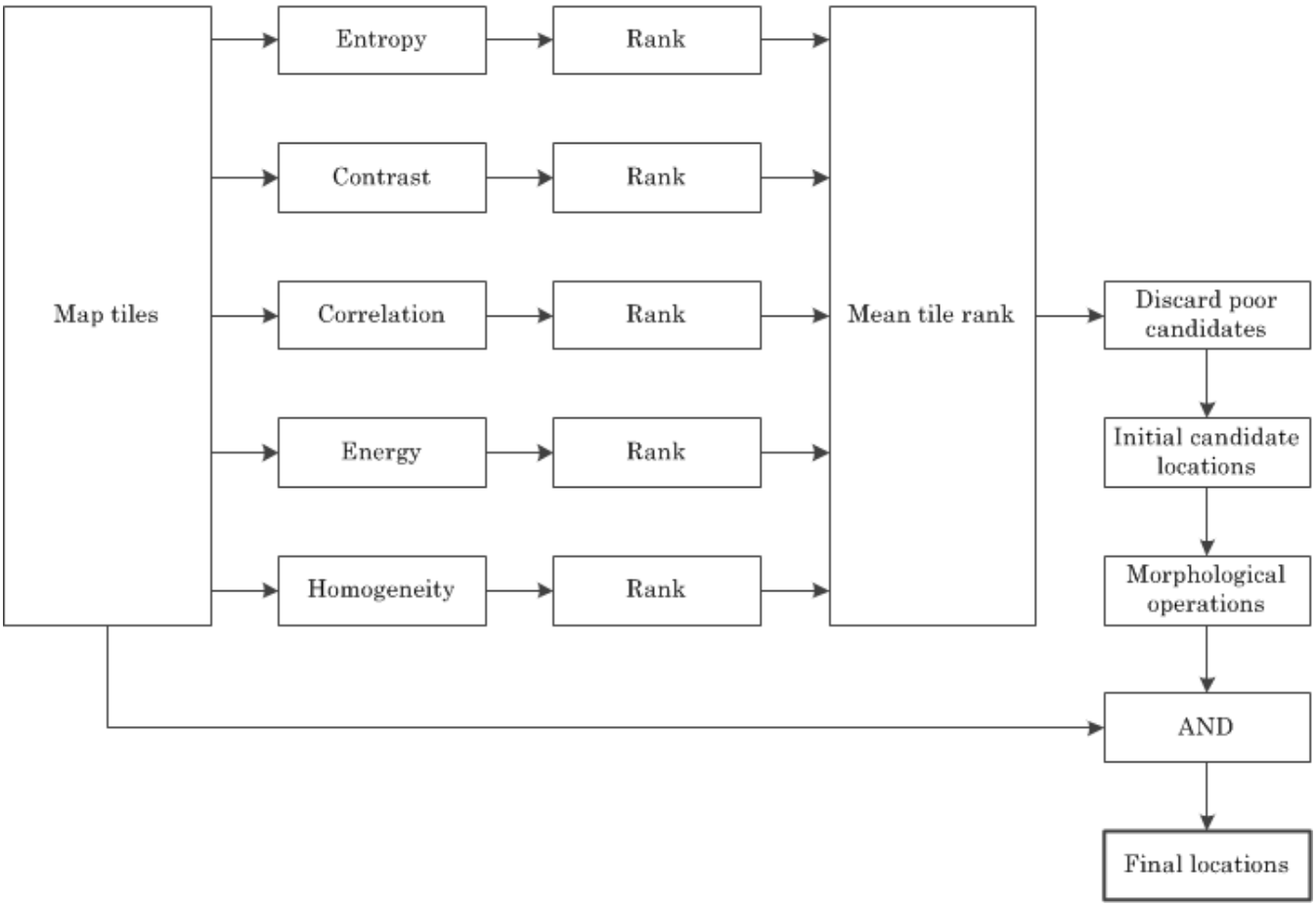


Figure 3: Process for determining obscured locations

The function $rank(t, metric)$ then corresponds to the particular rank (out of the set of all tiles and the same metric) of the scalar value for a given tile and texture metric. For each tile the arithmetic mean of its ranks across all texture metrics is then computed (Equation 7). Finally, the mean rank for each tile is normalized so that all scores are between 0 and 1.

$$MeanRank_t = \frac{\sum_{metric \in Metrics} s(t, metric)}{|Metrics|} \quad (7)$$

The result of this method is a consensus of the tiles with the lowest entropy and contrast and highest correlation, energy, and homogeneity. A portion of tiles with scores in the lower two thirds are discarded (masked), resulting in candidate locations (in the form of a mask that can be applied). This threshold (discarding tiles with scores in the lower two thirds) was determined through observation as a suitable and practical value and was applied consistently to all examples. However, such a threshold could be determined experimentally using Receiver Operating Characteristic curve analysis.

A series of morphological operations can then be applied to improve the mask and remove small, isolated regions. Erosion and then dilation were applied to the binary mask in order to produce the resulting region mask, shown in the right-most portions of Figure 4. Erosion thins objects, producing a smoother, smaller version; dilation enlarges the objects in the mask [10]. Future work could refine this method by investigating non-uniform weighting schemes for rank-level (or other) fusion of texture metrics for aerial imagery.

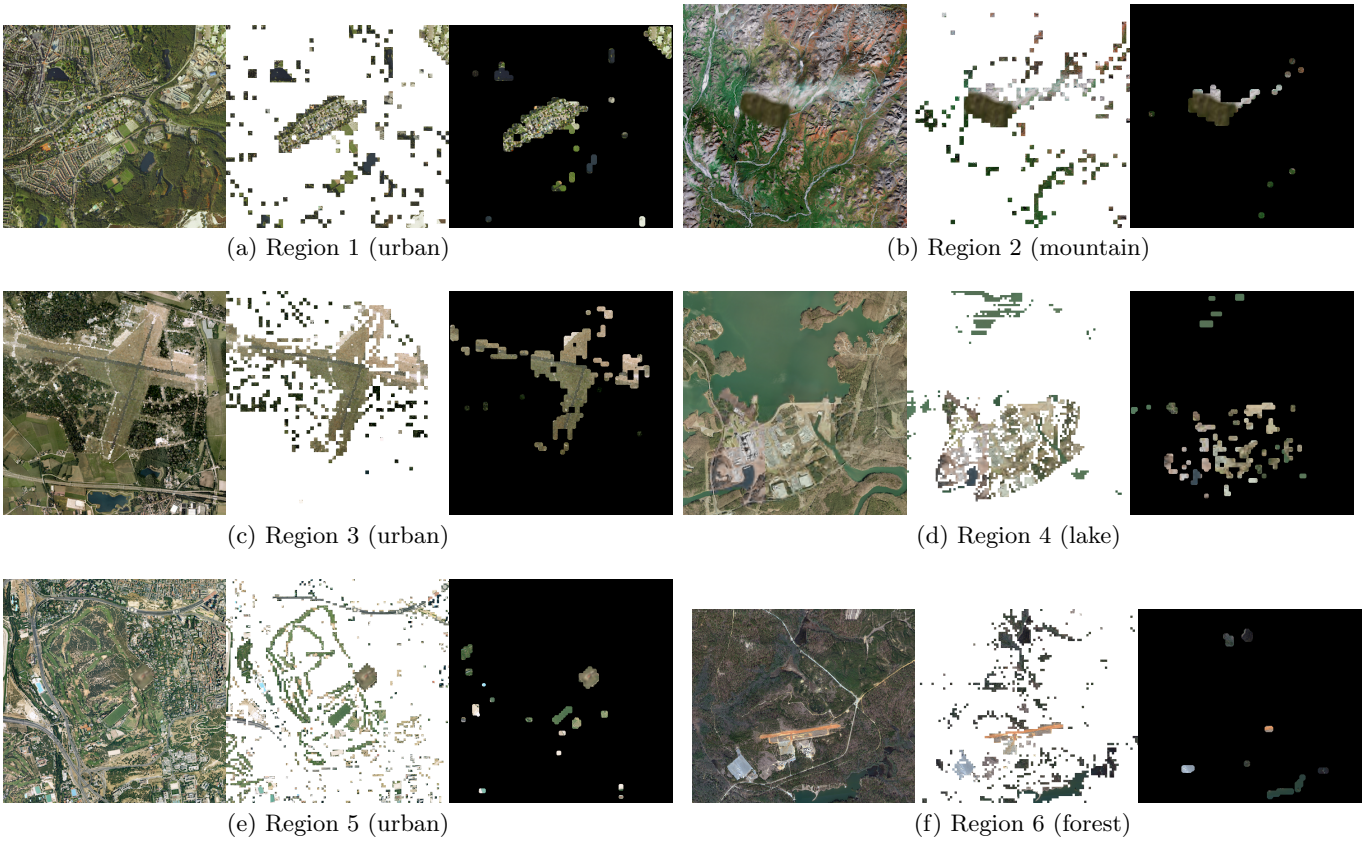
6.2 Results

The original and final, post-mask map areas are shown in Figure 4. In each example, the region of interest is generally preserved, although morphological operations are not able to preserve the area of interest in Region 6 (Figure 4f). Qualitatively, it can be observed that we successfully isolate the artificially-modified map data in five out of the six cases (although even one successful result would be sufficient to demonstrate the feasibility of this method). Further refinements could prune the potential distractors using more discriminant morphological operations.

This example demonstrated that detection of many map image manipulations is straightforward and can be automated. We believe that the plausibility of such an approach is clear. With more sophisticated experimental and analysis techniques, this approach can be extended to larger areas and a wider variety of obscuration techniques, while potentially further eliminating non-obscured map tiles.

6.3 Discussion

Taking all of the above a step or two further is trivial –



(a) Region 1 (urban)

(b) Region 2 (mountain)

(c) Region 3 (urban)

(d) Region 4 (lake)

(e) Region 5 (urban)

(f) Region 6 (forest)

Figure 4: Original map data (left), initial mask (center), and the final result after morphological operations are applied to the mask and combined with the original map data (right)

however, we have chosen not to publish any results of sites that deal with locations not mentioned in the Wikipedia list. Furthermore, the actually efficacy of our implementation of the technique is irrelevant to our conclusions. We are not attempting to prove that *our* detection technique works in all cases, but illustrate that the general approach works adequately. That the method works even for a single case (as has been presented here) is sufficient to broach discussion of the permanence of such security decisions.

In the highest level the recipe for discovering and exploring modified sites is truly trivial:

1. Crawl a source of images, such as Google Maps
2. Run the algorithm of your choice to determine locations which contain a high number of modified images
3. Remove (if you wish) "known" secret locations — what remains may be of considerable interest
4. Plug the GPS information in to other sources of image and look for differences manually
5. If the site looks interesting, plug the GPS coordinates into Foursquare, Flickr etc.
6. Look for restaurants and retailers near candidate locations — use sites like Yelp! to identify users who have reviewed local sites. Leverage the fact that sites often share similar user names.

The steps from here depend on the attacker's goal. Can this system be used in the manner described above? Absolutely! Will the technique provide actionable information? Yes. . . though we would hope that the sites discovered would actually be not too important. However, from the perspective of the person who chose to redact the image, the above technique is clearly not the desired outcome.

7. OLD PARADIGM, NEW WORLD

When governments chose to redact images obtained from satellites, it is possible they made a crucial mistake: by ignoring the probable trajectory of technology (multiple image sources, significantly faster computing, high bandwidth, cheap storage), they inadvertently highlighted those locations that were of importance to them without a corresponding gain in overall security. The techniques we have outlined above clearly show that it is possible for a relatively unskilled attacker to detect image manipulation techniques, and then use fairly low-tech and/or low-cost techniques to gather data. All the manipulation did, in effect, was tell attackers about locations that were of special interest.

Such issues – essentially, applying special protection to objects that are special – are hardly uncommon to the field of computer security. This approach fails when the labeling is not supported by sufficient protection. A classic example of this is found in the Orange Book, where B1 security requires the labeling of objects, but the requirement for their adequate protection is not enforced until B2 [3]. Similarly,

spotting a high value asset in physical security can be as simple as looking for the building with the best locks and encrypting only data which are important can allow for fairly easy traffic analysis by an adversary. Thus, we ask why the defenders chose the approach they did.

While we are unable to ask those who selected images for redaction, we have pondered the choices made. It is possible, for example, that the defenders were aware that the redaction would draw attention, but simply wanted to buy time to make some changes at the site. Similarly, the defenders may have not cared about additional attention, or used the redaction as a feint to draw attention away from more important (and unredacted) sites. Sadly, we cannot know, but at least to us, it seems likely that at least some redaction was designed to hide the actual redacted site.

That being the case, one possibility is that human nature is such that when a threat is present, we feel safer taking some action, even if it is ill-considered. Thus, there is a tendency to see the problem and hastily put in place a solution without thinking of the consequences carefully. The defender feels better because he is doing *something*.

Another possibility is that the defenders failed to understand the importance of the changed modality of the data. Document redaction is a common approach to maintaining secrecy: if a document contains sensitive information, a censor can simply cut out the parts which they deem dangerous. The problem is that the Google image data is fundamentally different to the release of a document: the ground truth (no pun intended) is available easily, and so the redaction only serves to highlight the issue. The data modality is important; for example, data sanitization is a tricky problem in network security experiments (see, for example [2]). Changing the IP addresses of just those flows which are sensitive would not pass as an acceptable sanitization process; imagery is no different. The change in the type of content (documents to real world imagery) perhaps helped drive the application of the wrong solution to a (subtly) different problem.

Overall, we believe the defenders made the following mistakes in their reasoning:

1. The defenders did not seem to be clear about what threat they were protecting from. Who was the attacker? What were their capabilities? Perhaps most importantly, what attack were the defenders attempting to stop?

2. The defenders did not seem to care that the security countermeasure raised the bar a negligible amount at the cost of drawing attention to the area of concern.

3. The defenders did not seem to anticipate the inevitable increase of availability of high resolution satellite imagery; the defensive countermeasure won nothing, at the cost of drawing attention to the very sites that were designed to remain incognito.

The need for this paper and argument was highlighted by an argument we have heard several times when discussing this research: "but that's clumsy manipulation... you can make more subtle changes that are much harder to find!" Alas, this response completely misses the thrust of our research. Despite the dangers inherent in image redaction and its associated downstream consequences, this espoused approach does not address the issue from the attacker's perspective. If the attacker knows the location of the site of interest, gathering non-obfuscated imagery is near trivial. If the attacker does not know the location of the site of interest, redacted the images can only serve to draw attention to the site, no matter how good the manipulation.

As geospatial data becomes more widely available, the risks posed by inadvertent leakage of such data grows. For example, customer reps who visit "secure" sites are often participants in services that allow users to track their movements indirectly. Leaving their cell phone outside the classified area provides no defense. Foursquare and geotagged photographs are but two examples of services which can allow an attacker to learn far more than we might at first predict; there are many more. A great deal can be learned from these services; knowing who supplies a site is valuable to a would-be attacker. Knowing precisely which person drives the delivery truck is even more so! Image manipulation opens the door to these attacks: it is much easier to find a needle in a haystack that it is to find a particular needle in a pile of other needles!

As we make security and technology decisions it is crucial that the permanence of data is recognized. That is, information does not go away: it is not enough to consider how the data can be leveraged now (when the US and the USSR had the only high-resolution imaging satellites, for example, the idea of manipulating satellite imagery data made sense – albeit barely) but how the data can be leveraged in the future. Security/disclosure decisions live long past their intended shelf life. A top secret site obscured ten years ago provides useful intelligence to adversaries today who can automatically detect the site.

These arguments are not limited to image data. Instead, in a world with near infinite memory, it is critical that security decisions are carefully weighed against a long term perspective. Some intelligence has a very short shelf life; some is valuable for many years. Discriminating the value of short-term hiding from long-term disclosure is important, as the growth of technology often renders preventatives ineffective after a certain amount of time. As noted above, this understanding is common in cryptography. If we are confident about the strength of an algorithm, defenders know approximately how long a secret will remain a secret when under attack. The key and algorithm are then chosen with this in mind. We argue urgently for the application of this technique to other areas.

The booklet "No More Secrets: National Security Strategies for a Transparent World" [9] argues that the traditional notions of secrecy are being erased by the interconnectedness of the modern world. The rate of production of information is incredible; this pool of data is only partly leveraged, and the role of Open Source Intelligence (OSINT) is likely to continue to grow over the next decade. This in itself adds to the complexity of making a good security decision, as the consequences of data release can be difficult to foresee, especially when coupled with other sources of data which the defender may be unaware of.

Essentially, most actions we take as defenders tell the attacker something; even adding a filter to a firewall may provide actionable information to an adversary. Furthermore, this information does not "go away" as a function of time. We believe that there are a class of problems where the release coupled with the long-term value of the information presents a problem. Above, we gave an extended example

from Google Maps. Encryption (especially steganography) is another area where the encrypted data, once released, can cause problems long after the disclosure. Another example is biometrics.

Steganography embeds a payload within a larger cover object. The intention is to hide the presence of the payload within a object that is safe for public transmission. However, the potential remains that the payload will later be detected if more sophisticated steganographic techniques are developed.

Biometric data is intrinsic. We cannot change who we are, nor can we practically alter characteristics such as our fingerprints or irides. However, these data can be assumed by others through various falsification techniques. Without proper handling method, exfiltrated biometric information can be used to gain unauthorized access to a resource. The consequences are lasting and more severe than exposing a password: a password can be changed or revoked, a biometric cannot. A poorly-designed biometric system may result in exposed credentials, threatening the security of other resources and users without a convenient method of invalidating those credentials.

7.1 The Way Forward?

In our discussions at the workshop, two points became much clearer to us. The first is that any time we distort reality (in essence, lie), we are entering a game where frequently the consequences of getting caught in our lie are more serious than revealing the underlying truth. The second point was perhaps more of a request: are there ways in which we can better reason about downstream consequences of security decisions. We discuss these two issues here.

Using distortion of reality to reveal a security problem is not new. The idea of a cross-view diff for detecting stealth software relies on imperfect camouflage — essentially, the ability to catch the attacker out in a "lie" [17]. For at least this detection technique, *only* the deception reveals the presence of the attacker. Had the attacker simply left stealth out of the equation, the attacker would not have been detected (though they would not have been hidden either). Many of these trade-offs can be captured in a simple game theoretic manner, comparing the probability and cost of choosing to lie and the consequences if detected. This approach has utility every time we choose to distort reality, and would have been a useful tool when attempting to reason about the help and risks in redacting online map imagery.

Other disciplines — especially management — frequently have to deal with a future which is unknown. Such strategic planning can be applied to security decisions once the need for this reasoning is recognized. Companies frequently have to peer into the future and attempt to manage risk. Security should be no different, though futures can be, perhaps, a little more unexpected when new exploit techniques are developed. At a minimum, security decisions that involve the distortion of reality and/or the release of information have strategic importance; it is important that procedures are put in place that recognize their difference from day-to-day tactical matters, and are treated accordingly.

8. CONCLUSION

In this paper, we have examined an example of a security decision that has had negative downstream consequences based on the evolution of technology. We argue that our tendency to deal with tactical issues (in this case, the ready availability of satellite imagery to non-government organizations) can lead to strategic missteps. Furthermore, in the case of imagery, our missteps have placed important information into the hands of attackers.

The paradigm here is not necessarily completely new, in that the best security researchers have made the argument before (and we follow happily in their footsteps), but it is clearly not adopted, as evidenced by the data we have gathered. Our hope is that by presenting the pointlessness (and indeed, in some cases, harm) of knee jerk reactions to threats we can learn to think in the long term, and add the question "how will foreseeable changes in technology impact this security decision?" to our analysis of security decisions made today.

While the geospatial genie is well and truly out of the bottle, it is not the only example of system design that needs to consider the impact of new technologies. As we illustrated, a similar issue is potentially brewing in the field of public key encryption, where the rapid developments in quantum computing pose a threat to traditional encryption techniques. Given the seeming inevitability of a usable quantum computer in the next N years, developers designing new systems should seriously consider using quantum resistant algorithms. The technology change is predictable, and the impact large; trying to address this once the technology is available is too late. We ask what other areas exist that we are not considering.

9. ACKNOWLEDGMENTS

The authors would like to thank the participants and reviewers of NSPW 2012 for their helpful and insighful comments. We especially thank the scribes who carefully noted the comments made during our presentation. These notes were invaluable in shaping the final manuscript.

10. REFERENCES

[1] M. Barbaro and T. Zeller. A face is exposed for aol searcher no. 4417749. *The New York Times*, Aug. 2006. Downloaded from http://www.nytimes.com/2006/08/09/technology/09aol.html.

[2] M. Bishop, J. Cummins, S. Peisert, A. Singh, B. Bhumiratana, D. Agarwal, D. Frincke, and M. Hogarth. Relationships and data sanitization: A study in scarlet. In *Proceedings of the 2010 Workshop on New Security Paradigms*, NSPW '10, pages 151–164, New York, NY, USA, 2010. ACM.

[3] Department of Defense. Department of Defense Trusted Computer System Evaluation Critera. DoD 5200.28-STD, Dec. 1985.

[4] H. Farid. Image forgery detection. *Signal Processing Magazine, IEEE*, 26(2):16 –25, march 2009.

[5] Flickr. Flickr website, April 2012. `http://www.flickr.com/map/`.

[6] G. Friedland, G. Maier, R. Sommer, and N. Weaver. Sherlock holmes' evil twin: on the impact of global inference for online privacy. In *Proceedings of the 2011 workshop on New security paradigms workshop*, NSPW '11, pages 105–114, New York, NY, USA, 2011. ACM.

[7] Geoeye. About us, April 2012. `http://geoeye.com/CorpSite/about-us/`.

[8] Google. Google maps website, April 2012. http://maps.google.com.

[9] A. Kochems. *No More Secrets: National Security Strategies for a Transparent World*. American Bar Association Standing Committee on Law and National Security, Jan. 2010.

[10] O. Marques. *Practical image and video processing using MATLAB*. Wiley Online Library, 2011.

[11] Microsoft. Microsoft bing maps website, April 2012. http://www.bing.com/maps/.

[12] OpenStreetMap. Openstreetmap website, April 2012. http://www.openstreetmap.org/.

[13] R. A. Perlner and D. A. Cooper. Quantum resistant public key cryptography: a survey. In *Proceedings of the 8th Symposium on Identity and Trust on the Internet*, IDtrust '09, pages 85–93, New York, NY, USA, 2009. ACM.

[14] C. Podilchuk and E. Delp. Digital watermarking: algorithms and applications. *Signal Processing Magazine, IEEE*, 18(4):33 –46, jul 2001.

[15] T. Reichhardt. The first photo from space. *Air & Space magazine*, Nov. 2006.

[16] Twitter. Twitter website, April 2012. http://twitter.com/.

[17] Y.-M. Wang, D. Beck, B. Vo, R. Roussev, and C. Verbowski. Detecting stealth software with strider ghostbuster. In *International Conference on Dependable Systems and Networks*, pages 368–377. IEEE, June 2005.

[18] Wikimapia. wikimapia website, April 2012. http://wikimapia.org/.

[19] Wikimedia. Geohack, April 2012. http://toolserver.org/ geohack/.

[20] Yahoo! Yahoo! maps website, April 2012. http://maps.yahoo.com/.