can execute active content, which in turn may be able to make calls and send multimedia messages. If active content is able to create active content to other devices then, of course, self-replication is possible and thus viruses infecting mobile devices are possible. Until now, mobiles phones have been closed environments, but that is changing. We do not know what will happen after the Cabir virus, but what we do know is that current mobile technology allows viruses to exist in mobile devices. What is frightening is that mobile phones are likely to have more and more complexity, features, and capacity."

Now that we know that these kinds of devices can be infected by malicious code what should we do? Leave Bluetooth disabled unless you really need it because apart from the mobile phone virus, which is not in the wild, Bluesnarfing has arrived. "For Bluesnarfing to be successful", says Colm Murphy, "the sender and the recipients need Bluetooth enabled on their mobile phones. You simply search for all the Bluetooth enabled devices within a 30-feet radius and send your message. This could be a derogatory remark, a marketing ploy at an Expo– 'free coffee and cakes at stand 51', or anything your imagination feels free to conjure up. Things could easily get out of hand with this facility, especially in relation to bullying or sexual harassment. One way to avoid this and take yourself out of the loop is to set your phone to only accept or send messages from and to a preferred list, or simply disable Bluetooth", says Murphy.

"I think the 'Expo' example is a good one", says Murphy. "It is perfect ground to release a mobile virus that spreads quickly and targets a specific audience. Or what about a mobile phone virus that makes the phone dial a specific number! Some disgruntled ex-employees with a grudge could have some fun with that one!"

"The big concern is that in the future large mobile virus outbreaks may be reality", says Dr. Helenius. "Denial-of-service attacks may affect critical infrastructures, like emergency phone numbers. Indeed, an efficient virus may be able to block phone lines and phone networks, like an efficient Internet worm can block Internet connections. However, we do not know if such disasters will happen. The mobile device and network developers have a choice. They can adapt more security in their products in order to prevent disasters. Software could be written securely in order to prevent buffer overflows and other critical errors. More importantly, security can be an essential part of design. For example, it is possible to adapt hardware components that will prevent unauthorized phone calls. We should not merely trust software, and security should be based on more than one layer. If one security layer fails, there could be other security layers that will prevent further damage." On a more skeptical note Colm Murphy says, "If you consider the tens of millions of EURO that companies are investing in developing products that stop these kinds of threats, I think it is safe to assume that we will see more mobile phone viruses as time passes. The investors will demand it!!!"

References:

[1] It spreads to devices that run under Symbian OS, which is used in many models of phones manufactured by Nokia, Siemens, Sony and Ericcson.

[2] Even a Bluetooth-enabled printer according to the Symantec security response.

[3] (www.symantec.com)

IDS or IPS: what is best?

Maria Papadaki and Steven Furnell University of Plymouth,

Intrusion detection systems (IDS) have become one of the most common countermeasures in the network security arsenal. But while other technologies such as firewalls and anti-virus provide proactive protection, most current IDSs are passive; detection of a suspected intrusion typically triggers a manual response from a system administrator. Too often, this comes too late.

Cohen has demonstrated the importance of quick response. In a simulation study he showed that if 10 hours elapse between detection and response, then attackers have an 80% chance of success. At 20 hours, the success rate rises to 95%, and after 30 hours the attacker will always succeed, regardless of the skill of the administrator. But if the response is instant, the probability of success against a skilled administrator is almost zero. [1]

Such findings are valuable in view of the rapid escalation that characterises many of today's Internet-based attacks. Recent incidents such as Sasser and MyDoom have shown us that we do not have the luxury of time to react.

Such factors have led to increased interest in an alternative technology, namely the intrusion prevention system (IPS). Although these incorporate intrusion detection mechanisms, and share similarities such as being deployable in both network and host-based contexts, they also have two significant differences. Instead of passively monitoring activity on systems and networks, IPSs are positioned inline and can therefore block unauthorized activity before it takes place (see Figure 1). In a network context, conceptually they combine firewall approaches with intrusion detection capabilities; in host environments, they monitor all system and API calls and block those that would cause malicious behaviour. [2]

With reference to Figure 1, products are now available that can be configured to operate in either mode (an example being McAfee's Intrushield). However, as later discussion will establish, this is not to suggest that the use of IDS and IPS is an either/or decision.

The IDS is dead, long live the IPS?

Although IPS solutions have been available for several years, their adoption had been limited. More recently, however, there has been a shift in the attitudes of vendors and users in relation to IDS, and in the characteristics of the products on offer.

A notable contributor to this was a market report from Gartner in June 2003. This set tongues wagging because it branded IDS technology a "market failure", and predicted that it would be dead by 2005.[3] The report suggested that customers hold off big investments in IDS because the technologies added no practical value in enterprise security. Reaction to this report from IDS vendors and security specialists was intense.[4]

Gartner argued against IDS mainly because of their inability to prevent intrusions, and the vast number of false positive alarms they can generate. False alarms are indeed a recognized problem with IDS, and are the bane of many security administrators' lives. A 2003 survey by OpenService, Inc (www.open.com) established that management of false positives is among the top three problems facing security practitioners; only shrinking budgets and threat risk assessments raised more concern. [5]

The tendency of IDS to generate false positives also has the undesirable side-effect that administrators tire of following-up dead ends and become slack about tracking fresh alerts.

Gartner's other main point, that IDS does not prevent intrusions, is also fair. Usually this is because the IDS is placed out of band as a monitoring device, with its response capability restricted to passive actions such as logging data and issuing alerts. Given the problem of false positives, it is understandable that IDS is not often trusted to respond more actively, such as blocking traffic, ending sessions, restricting access and the like.

The debate had forced many IDS vendors to incorporate intrusion prevention solutions in their products. Even where vendors have not adopted preventi on solutions, the term "intrusion detection system" tends to be avoided. In its place people talk of "intrusion management system" or "intrusion protection system". This may be to distance products from any doubts in potential customers' minds. Indeed, some effects amongst the user community can also be observed. For example, for the first time, the CSI/FBI security survey reports fewer respondents using IDS technology (see Figure 2). It is also notable that the 2004 survey was the first to ask respondents specifically about the use of IPS technology. The question got a 45% response rate.[6]

So, why the sudden interest in IPS products? Is it vendors running scared, wanting to distance themselves from the fallout from the Gartner report? Is it a marketing exercise? Refocusing of products certainly has the potential to press the right buttons from a consumer perspective—after all, why would you want to buy a detection product if you can get one that actually prevents intrusions? Or have we hit upon a technology that solves the problem of attacks, without the perceived weaknesses of IPS? In short . . .

Is IPS really an alternative?

The ability to stop intrusions logically suggests a maturing of the technology. It suggests that intrusion detection technologies have become accurate enough for us to rely upon their decisions to be correct. Without the IDS-related concerns over accuracy and false positives, we can thus rely on them to issue preventative responses with confidence. Unfortunately, however, intrusion prevention systems do not have a silver bullet for this problem; they may in fact use the same detection methods as IDS.

The solutions provided by IPS products therefore attempt to sidestep the problem of false positives by only blocking those attacks that can be detected with high certainty. In effect, this means transfer of the strongest and most reliable technologies from the IDS domain into a different mode of operation. Even then, IPS cannot be regarded as a fix of the problem of false alarms. The solutions will seldom work perfectly "out of the box"; most require tuning to tailor their most effective operation.[7]

The biggest advantage of intrusion prevention is its potential to respond in real time and to nip attacks in the bud. However, as promising as it sounds, there are concerns about the IPS approach. The first is the overhead they can introduce in networks and systems by having to authorize all traffic and all system calls. This becomes more significant in busy

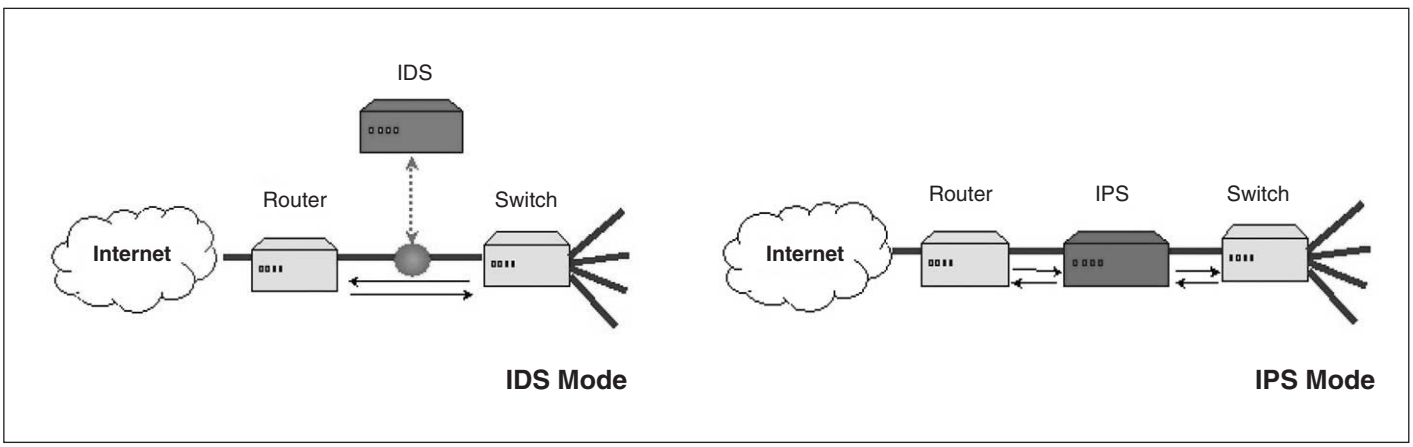

Figure 1: Offline vs. inline placement

networks and servers, where performance is crucial.[8] At the same time, however, there are parallels with firewall technologies, and devices can be designed and deployed with performance considerations in mind.

Single point of failure

A potentially worse problem is that IPSs are a single point of failure. An error here could have significant impact upon systems and networks. For example, if an IPS crashed because it couldn't cope with the traffic, or was the target of an attack, the disturbance on the network's operation would be considerable. There are some moves to overcome this problem (e.g. using a back-up IPS that takes over in an emergency, or reconfiguring the router to redirect traffic around the IPS, or pre-configuring the IPS to run with minimum capabilities, allowing all traffic to pass), these solutions do not fully address the issue because systems may be unprotected.

The problem of volume-related crashes is well known. Vendors are improving their products, but still have issues to address. So far, a good solution remains elusive.[9]

A more significant concern is to avoid false positives. Killing only the most suspect attacks means that a range of different attacks may pass because the cautious IPS does not recognize them as intrusions. In this scenario, a further line of defence is still very desirable.

So, to answer the question posed at the head of this section, IPS is not an alternative to IDS; it is not meant to be. But the technologly does provide another layer of security, which is important in a defence in depth strategy. As such, both IDS and IPS have important roles, and it should not be the case that one is used in place of the other.

Improving our response

But it is essential that either approach provide a correct response to the intrusion. Indeed, the fact of IPS and its attractiveness is closely linked to the need to respond.

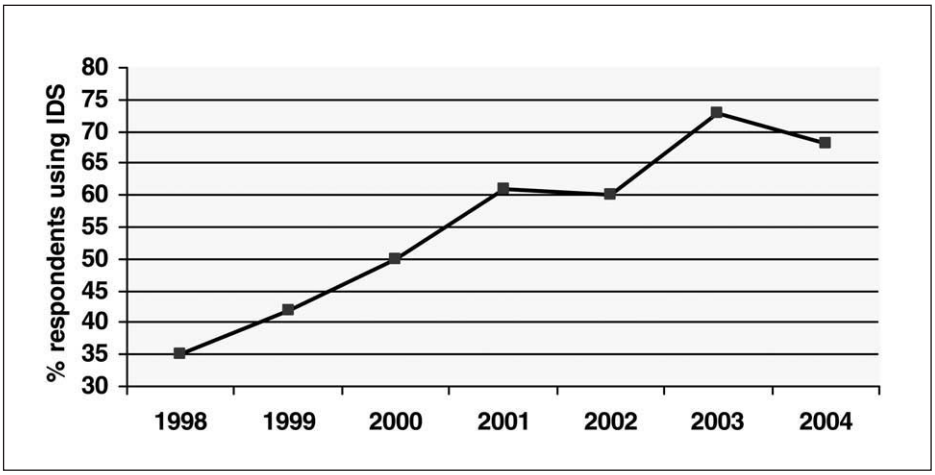

Figure 2 : Organisations using IDS technology (source: CSI/FBI surveys)

However, blocking sessions and dropping packets is not the only appropriate response. It should be possible to limit more subtle attacks by investigating them further to reach a more informed judgement.

Indeed, an IPS configured with a simple "block or pass" strategy may not provide enough flexibility; the fact that they respond only to the most definite signatures means that false negatives can occur. As such, it would be unwise to consider IPS as the only defence against intrusions. It therefore makes sense to subject the traffic that gets through to further analysis using an IDS.

At this point, however, the question still remains about what the IDS should actually do if it finds something it believes to be intrusive. If the technology is simply used in its current form, then the need to limit responses to passive actions for fear of false positives will be largely unchanged.

An alternative strategy is to endow the IDS with a sense of its own inadequacy. In other words, account for the fact that some detection judgements are likely to be stronger than others, and allow flexible levels of response to be issued accordingly.

Our group and others have researched this. Given that today's commercial IDS technologies are rooted in research from the 1980s, we should consider how current research may help to advance the commercial incarnations in the future. Our results suggests that incorporating two aspects is particularly desirable:

- Adaptation of responses according to the current context.
- Assessment of the appropriateness of response actions before and after initiating them.

Adaptive decision-making relates to the need for response decisions to change with the context in which an incident has occurred (i.e. a response that is appropriate to a one type of incident on one occasion will not necessarily be appropriate if the same incident happened again under different circumstances).

When considered in terms of two collaborating IDS entities, a Detection Engine and a Responder, the determination of this context may include a number of considerations. These include

- Whether the incident is part of an ongoing series of attacks (e.g. how many targets have already been affected? Which responses have already been issued?) .
- The current status of the target (e.g. is it a business critical system? What is its load at the moment? Is there any information or service that needs to be protected? What software/hardware can be used for response?).
- The perpetrator of the attack (is there enough information to suggest a specific attacker? Is he/she an insider/outsider?).
- The privileges of the user account involved (e.g. what is the risk of damage to the system?).
- The probability of a false alarm (how reliable has the sensor/source that

detected the incident been in the past? What is the level of confidence indicated by the Detection Engine about the occurrence of an intrusion?)
- the probability of a wrong decision (how effective has the Responder been so far? Have these responses been applied before in similar circumstances?).

Having assessed the above factors, response decisions must then be adapted to the context accordingly. For example, if the incident has been detected on a business critical system, but the Detection Engine has indicated a low confidence, then the selection of a response with minimal impact upon the system would represent the most sensible course of action (i.e. minimizing the chance of critical operations being disrupted in the case of a false positive). However, if the same scenario occurred in conjunction with previous alerts (i.e. showing that the current incident is part of a series of attacks), then a more severe response is warranted.

The other required feature is the ability to assess the appropriateness of response actions. There are two ways to do this, firstly by considering the potential side effects of a response action before issuing it, and secondly by retrospectively analyzing its effectiveness in containing or combating attacks.

The problem of side effects is a particular concern when using active responses (e.g. blocking, termination and access restrictions) because they may disrupt legitimate users. As a result, the response needs to be considered before a given action is executed. Several characteristics would be relevant to consider in this context:

- The transparency of the response action. In some cases it might be preferable to issue responses that do not alert the attacker to the fact that he/she has been noticed; in others it could be preferable to respond very explicitly.
- The degree to which the action would disrupt the user against whom it is issued. This is especially relevant when a response is mistakenly issued against a legitimate user. In situations where the Detection Engine has flagged an incident, but expressed low confidence, it may be desirable to start by issuing responses that a legitimate user would be able to overcome easily.
- The degree to which the action would disrupt other users, or the operation of the system in general. Certain responses (e.g. termination of a process, restriction of network connectivity) would affect more than just the perceived attacker. As such, the Response Policy may wish to reserve such responses only for the most extreme conditions.

Each of these factors needs to be assessed independently, and incorporated into the response selection process as appropriate, as well as during the formulation of the Response Policy by the system administrator.

The second factor that would influence appropriateness is whether a suspected attack has been used before in the same context. If the Responder keeps track of its previous response decisions, then they can be used later to assess whether the selected actions were actually effective. This requires a feedback mechanism that can then be used to refine the Response Policy.

Feedback could be provided in two ways: explicitly by a system administrator, and implicitly by the Responder itself. In the former, the administrator would inspect the alert history and manually provide feedback in relation to the responses that had been selected. This would say whether or not they had been effective or appropriate to the incident.

Otherwise, the Responder itself could infer whether previous responses had been effective. For instance, it could say whether it had been required to issue repeated responses in relation to the same detected incident. If this was true, it could potentially infer that (a) the initial response actions were not effective against that type of incident, and (b) the last response action issued might form a better starting point on future occasions (i.e. upgrading and downgrading the perceived effectiveness of the responses when used in that context).

We have a prototype implementation of the above approach. It forms part of PhD research work within our group.[12] It has provided a practical proof of concept for the ideas expressed here, and suggests that the long-term choice in the intrusion-handling domain may be broader than current detection and prevention technologies.

Having said this, we need to do some more work on it before it is ready for large-scale deployment.

Conclusion

The title of this article was perhaps a little misleading, in the sense that neither technology is a complete answer. Although IPS technologies provide a way to thwart high-certainty attacks, we still need the IDS to account for other cases.

This leaves us with a problem. The imperfect nature of detection means that we can easily mistake normal activity for an intrusion. At the same time manual responses could be too late to prevent incidents.

We advocate a more flexible and intelligent approach, one that offers escalating levels of response according to several contextual factors. Although we have worked on this with some effect, we need to do more to reduce the uncertainty in response decisions before we look to automate fully prevention and response activities.

About the authors

Maria Papadaki has recently completed her PhD research within the Network Research Group, focusing upon the issue of flexible, automated IDS response. *This research activity was undertaken with support from the State Scholarship Foundation of Greece.*

Dr Steven Furnell is head of the Network Research Group at the University of Plymouth, UK. He is author of "Cybercrime: Vandalizing the Information Society", published by Addison Wesley.

References:

[1] Cohen F.B. 1999. "Simulating Cyber Attacks, Defences, and Consequences", The Infosec Technical Baseline studies, March 1999. http://all.net/journal/ntb/simulate/simulate.html.

[2] Network Associates. 2003. The Path to Prevention, White Paper, Network Associates Technology, Inc, October 2003. http://www.nai.com/

[3] Gartner. 2003. "Gartner Information Security Hype Cycle Declares Intrusion Detection Systems a Market Failure", Gartner Press Release, 11 June 2003.

[4] Taylor S. and Wexler J. (2003) "IDS vs. IPS: Is one strategy 'better?'", Network World Fusion, 16 October 2003. http://www.nwfusion.com/newsletters/frame/2003/1013wan2.html

[5] OpenService, Inc. 2003 Security Event Management Survey Results Analysis: Insight into the Threats, Issues and Trends Facing Network Security Departments in 2003. White Paper. February 2003. www.open.com.

[6] Gordon, L.A., Loeb, M.P., Lucyshyn, W. and Richardson, R. 2004. Ninth Annual CSI/FBI Computer Crime and Security Survey. Computer Security Institute.

[7] McAffee Security. 2003. Intrusion Prevention: Myths, Challenges, and Requirements. White Paper. Network Associates. April 2003.

[8] Messmer, E. 2002. "Intrusion prevention' raises hopes, concerns", Network World Fusion, 4 November 2002, http://www.nwfusion.com/news/2002/1104prevention.html

[9] Snyder, J. 2003. "False positives remain a major problem", Network World Fusion Online Magazine, 13 October 2003.http://www.nwfusion.com/reviews/2003/1013idsalert.html

[10] Papadaki M., Furnell S.M., Lines B.M., and Reynolds P.L. 2003. "Operational Characteristics of an Automated Intrusion Response System", in Communications and Multimedia Security: Advanced Techniques for Network and Data Protection Lioy A. and Mazzochi D. (eds), Springer Verlang, October 2003: pp 65-75.

[11] Ragsdale, J.D., Carver, C.A. Jr., Humphries, J.W., and Pooch, U.W. 2001. "Adaptation Techniques for Intrusion Detection and Intrusion Response Systems", 2nd Annual IEEE Systems, Man, and Cybernetics Information Assurance and Security Workshop, West Point New York, June 5-6 2001.

[12] Papadaki, M. 2004. Classifying and Responding to Network Intrusions. PhD Thesis. University of Plymouth, United Kingdom.

The big picture on big holes

Thomas Kristensen, Secunia

Broken Zones in Internet Explorer

The Security Zone model of Internet Explorer has been criticised almost continuously, and once again it has been compromised. This time it appears that a Web master has used it to maliciously to install adware.

Internet Explorer has had multiple inappropriate functionalities and minor security issues for a long time. Alone, these issues doesn't pose a real threat to Internet Explorer users; however, combined with other vulnerabilities that allow websites to interact with the Local Security Zone, it is possible to place and execute arbitrary code on users' systems with no warning or user interaction.

Currently, there is only one effective solution: disable Active Scripting. However, most businesses and private individuals find this is inappropriate, as too many sites rely on Active Scripting to work properly.

Many professionals believe there is a better solution: use another browser. Even US-CERT/cert.org has suggested (recommended) use of an alternative browser, a solution rarely suggested by US-CERT.

We can only hope that Microsoft issues an out of schedule patch to deal with these latest issues. Hopefully it will soon release a cumulative patch to remove some of the inappropriate features and minor security issues that permit someone to compromise a full system.

Spoofing IE

Yet another spoofing vulnerability was found in Internet Explorer. This allows malicious people to change the appearance of a domain in the address bar. This could be used to establish outbound SMB/CIFS connections, if this kind of traffic isn't properly filtered at the perimeter firewalls.

Internet Explorer Jelmer issued a detailed analysis of a very sophisticated "zero-day" exploit for Internet Explorer. Jelmer obtained the exploit from an adware site that is using this exploit to install a toolbar in Internet Explorer on vulnerable users' systems.

Please read Secunia advisory SA11793 below for additional details.

Furthermore, Microsoft's monthly security bulletins for June addressed vulnerabilities in DirectX and various products that implement Crystal Reports.

http://secunia.com/SA11793
http://secunia.com/SA11803
http://secunia.com/SA11802

Another vulnerability was identified in Internet Explorer. This could be exploited by a malicious website to bypass security zone restrictions and spoof the address bar. Only Mozilla suffers the same vulnerability. However, in Mozilla's